

汉语小句的俄语对应单位研究*

杨毅¹, 冯文贺^{2✉}

(1. 乌拉尔联邦大学 计算机词典编纂问题实验室, 俄罗斯 叶卡捷琳堡 620000; 2. 广东外语外贸大学 语言工程与计算实验室, 广东 广州 510006)

摘要: 标注汉俄平行文本中汉语小句的俄语对应单位, 并统计分析。首先, 根据汉语小句切分对齐切分俄语, 得到俄语对应单位。其次, 对俄语对应单位进行语法标注。最后, 基于标注语料, 分析发现, 俄语对应单位: 1) 句子组成部分多 (74.85%), 句子少 (25.15%); 2) 单一述谓核心多 (69.04%), 无述谓核心其次 (27.63%), 多述谓核心少 (3.33%); 3) 单一述谓核心以简单谓语最多 (31.84%), 无述谓核心以动词短语最多 (51.26%), 多述谓核心以主从复合句最多 (47.92%)。

关键词: 小句对齐; 篇章翻译; 汉语小句俄语对应单位

中图分类号: TP391

文献标识码: A

Study on the Russian Syntactic Analogues of Chinese Clauses

YANG Yi¹, FENG Wenhe^{2✉}

(1. Laboratory of Computer Lexicography, Ural Federal University, Ekaterinburg 620000, Russia;
2. Laboratory of Language engineering and computing, Guangdong University of Foreign Studies, Guangzhou, Guangdong 510006, China)

Abstract: This paper is devoted to annotate and analyze the Russian syntactic analogues of Chinese clauses based on the Chinese-Russian parallel official documents. Firstly, segment the original text (Chinese) into clauses, and segment target text (Russian) into corresponding Russian syntactic analogues based on the segmentation of Chinese clauses. Then, we set the general principles and annotation system to annotate these Russian syntactic analogues. Finally, we examine the following results: 1) the components of sentences (74.85%) are more than the sentences (25.15%). 2) Single predicative center accounted for the largest proportion (69.04%), which are more than non-predicative center (27.63%) and multi-predicative center (3.33%). 3) The simple predicate accounted for the vast majority of the single predicative center (31.84%), the verb phrases accounted for the vast majority of the non-predicative center (51.26%), and the subordinate sentences accounted for the vast majority of the multi-predicative center (47.92%).

Key words: clause alignment ; discourse translation; Russian syntactic analogue of Chinese clause

1 引言

小句 (clause, клауза) 是篇章翻译的基本单位。研究表明, 在机器翻译系统中, 基于小句的翻译比基于大句 (Sentence) 的翻译, 翻译结果的可接受率高出 45%^[1]。由此, 基于小句的翻译模型也成为篇章机器翻译研究的重要课题^[2]。统计翻译的根基在于有充分的双语对齐样本和双语语法知识, 相应地, 基于小句的双语对齐语料库与语法标注库的研制与分析就成为面向篇章机器翻译的基础性重要课题。

汉俄翻译中, 汉语小句的俄语对应单位复杂多样, 选择何种对应单位是汉语小句俄译的关键问题。如例 (1) 中, 汉语小句 C1-C5 的俄语对应单位分别为 R1-R5。其中 R1 为副动词短语, R2 和 R4 是简单句组成部分、R3 是前置词短语, R5 是主从复合句。

* 收稿日期:

定稿日期:

基金项目: 中国国家留学基金 (201508090048); Программы повышения конкурентоспособности Уральского федерального университета(02.A03.21.0006); 教育部人文社科项目 (13YJC740022、15YJC740021、16YJC740053); 河南高校哲社基础研究重大项目 (2015-JCZD-022); 国家自然科学基金 (61273320); 中国博士后基金 (2013M540594); 广东外语外贸大学语言工程与计算机实验室拓标课题 (LEC2016ZBK T001)

作者简介: 杨毅 (1989—), 女, 博士在读, 理论语言学, 语料库语言学; 冯文贺 (1976—), 通讯作者, 男, 博士, 主要研究理论语言学、计算语言学。

(1)^{C1} 保持经济中高速增长,^{C2} 推动产业迈向中高端水平。^{C3} 实现全面建成小康社会目标,^{C4} 到 2020 年国内生产总值和城乡居民人均收入比 2010 年翻一番,^{C5}“十三五”时期经济年均增长保持在 6.5% 以上。

^{R1}Поддерживая умеренно высокие темпы экономического роста, ^{R2}стимулировать выход всех отраслей экономики на средний и высокий уровень развития. ^{R3}Для достижения намеченных целей полного построения среднезажиточного общества ^{R4}нужно удвоить к 2020 году ВВП и среднедушевые доходы городского и сельского населения по сравнению с 2010 годом. ^{R5}Это значит, что среднегодовой прирост китайской экономики в период 13-й пятилетки должен быть не ниже 6,5%.

如例(1)所示,如能提供汉俄小句对齐样本,并进一步给出对应小句的语法知识,将为双语小句转换提供非常有效的知识指导,也将为机器翻译等研究奠定基础。至今,汉俄平行语料库已有很多^[3-5],但这些平行语料库一般基于操作性强的句子(一般以句号等为标志)甚至段落对齐,而基于小句对齐的平行语料库仅开始初步实践^[6-7],汉俄小句语法知识的标注资源几乎没有。

本文标注汉语小句与其俄语对应单位,并进行统计分析。基本步骤如下:首先,在汉俄平行文本上,以一定标准切分出汉语小句,以汉语小句切分为依据,对齐切分出俄语对应单位,形成汉俄小句对齐语料库(第2部分);其次,制定俄语对应单位的语法分析原则及体系,实现汉俄小句语法标注语料库(第3部分);最后,基于标注语料库分析俄语对应单位的统计分布(第4部分)。

2 汉俄小句对齐切分

本文汉俄小句对齐切分主要采用冯文贺等在创建汉英小句对齐语料库中^[8]提出的方法及原则,即:首先,依据一定标准切分汉语小句;其次,以汉语小句切分为依据,按照语义对应原则对齐切分俄语,得到汉语小句的俄语对应单位。

汉语研究对小句(*clause, клауза*)至今没有统一的定义与切分标准。本文采用李艳翠等^[9]对汉语“子句”有如下定义:“子句是篇章分析的基本单位,含传统单句和复句中的分句。结构上,子句至少包含一个谓语部分,至少表达一个命题;功能上,子句对外不作为其他子句结构的语法成分,子句和子句间发生命题关系;形式上,子句间一定有标点分割,通常是逗号、分号和句号等。实际语料中,一些与典型子句在结构、功能、形式上类似的传统所谓短语在特定条件下也作为子句处理。”该标准已被应用于建立汉语篇章结构语料库^[10]和汉英篇章结构语料库^[11]及自动分析研究,并取得较好效果。

以汉语小句切分为依据,在俄语翻译文本中得到语义上与汉语小句对应的俄语篇章单位。例如在(2)中,汉语切分为 C1-C5 共五个小句,语义上对应的俄语对应单位切分为 R1-R5。在语法上,R1、R2、R3、R5 是典型小句,R4 则是句组。如按照俄语本身性质,R4 可以进一步分为两个小句(“стремились к тому, чтобы”和“проводились”)。由此可见,汉语小句的俄语对应单位切分并非是按照俄语小句标准切分而来,而是以汉语小句为标准的切分。

(2)^{C1} 我们实施积极的财政政策和稳健的货币政策, // ^{C2} 坚持正确处理保持经济平稳较快发展、调整经济结构和管理通胀预期的关系, / ^{C3} 更加注重把握好政策实施的重点、力度和节奏, // ^{C4} 努力做到调控审慎灵活、适时适度, / ^{C5} 不断提高政策的针对性、灵活性和前瞻性。

^{R1}Мы проводили активную финансовую и уравновешенную монетарную политику, // ^{R2}продолжали правильно разрешать отношения между поддержанием ровного и относительно быстрого развития экономики, урегулированием экономической структуры и управлением ожидаемой инфляцией. / ^{R3}Придавали большее значение правильному

определению главных участков реализации политики, степени ее воздействия и ритмичности осуществления, // ^{R4}стремились к тому, чтобы регулирование и контроль были взвешенными и гибкими, проводились в подходящий момент и в разумной мере. / ^{R5}Непрерывно повышали целенаправленность политики, ее гибкость и опережаемость.

基于以上切分原则, 选取 2016 年中国政府工作报告的汉俄双语文本, 进行汉俄小句切分。以下对俄语对应单位进行句法标注。

3 俄语对应单位的语法分析

根据一定分析原则和语法体系对俄语对应单位进行语法分析与标注, 形成俄语对应单位的语法标注库。

3.1 分析原则

实践表明, 按照汉语小句切分的俄语对应单位并不全都是典型句法单位, 对此, 俄语对应单位语法单位性质的认定需要重新制定分析原则。

第一, 认定对象主体以整体为主, 局部为辅, 整体优先局部为原则。与汉语小句对应的俄语对应单位结构复杂, 包含多种句法结构。如例(3)中, 结构上, R1 是简单句, R2 是副动词短语, R3 为主从复合句; 整体上, R1-R3 为并列复合句。该部分的判定以其所在整体的句法性质而定。所以, R1、R2 和 R3 被认定为并列复合句组成部分。

(3) ^{C1}新修订的环境保护法必须严格执行, // ^{C2}对超排偷排者必须严厉打击, / ^{C3}对姑息纵容者必须严肃追究。

^{R1}Необходимо строго соблюдать пересмотренный Закон об охране окружающей среды, // ^{R2}сурово карая согласно закону за превышение нормативов предельно допустимых выбросов и нелегальные выбросы, / ^{R3}а также строго спрашивая по закону с тех, кто попустительствует этим правонарушениям.

第二, 以汉语小句为依据切分的俄语对应单位中有时会出现省略现象而影响到该语法性质的判断, 对其性质认定以将省略部分补充完整后的整体为对象分析。如例(2)中, R2 省略了主语(мы), 分析中应补出, 并认定 R2 为双部句。

第三, 带有同等谓语句子在俄语语法研究中一直存有争议, 其争议的焦点是这类句子应当视为简单句还是复合句^[12]。同等谓语分为名词性合成谓语和动词性合成谓语。在标注过程中, 我们把动词性合成谓语中的同等谓语部分标注为动词短语, 如例(4)中, R1、R2、R3、R4 为主从复合句组成部分, R1 为无人称句, R2、R3 为作为 R1 同等谓语的动词短语。

(4) ^{C1}必须毫不动摇坚持以经济建设为中心, / ^{C2}推动科学发展, / ^{C3}妥善应对挑战, // ^{C4}使中国经济这艘巨轮破浪远航。

^{R1}Нам необходимо без малейших колебаний отводить экономическому строительству центральное место, // ^{R2}стимулировать научное развитие, // ^{R3}должным образом реагировать на риски и вызовы с тем, / ^{R4}чтобы обеспечивать китайской экономике, подобной гигантскому кораблю, дальнейшее плавание вперед навстречу волнам.

对名词性合成谓语的认定则将省略部分补充完整后分析, 这是由于名词性合成谓语虽无时态的变化, 却有数和人称的变化, 见例(5)。

(5) ^{C1}取消和下放 311 项行政审批事项, ^{C2}取消 123 项职业资格许可和认定事项, ^{C3}彻底终结了非行政许可审批。

^{R1}Были отменены либо делегированы вниз полномочия по 311 пунктам административного утверждения, ^{R2}аннулированы 123 пункта лицензирования и сертификации профессиональных квалификаций, ^{R3}окончательно отменено неадминистративное утверждение.

3.2 语法分析体系

基于俄语传统句法理论对俄语对应单位进行分析,结合语料分析实践,形成俄语对应单位的语法性质体系。

首先,根据句法完整性,把俄语对应单位分为:句子(简单句,并列复合句,主从复合句,无连接词句,句组)和句子组成部分(简单句组成部分,并列复合句组成部分,主从复合句组成部分,无连接词句组成部分)。

其次,根据是否包含述谓核心,把俄语对应单位分为:无述谓核心、单一述谓核心和多述谓核心。无述谓核心包括名词短语,动词短语,前置词短语,副词短语,副动词短语和形动词短语;单一述谓核心包括单部句(无人称句、不定式句、确定人称句)和双部句(简单谓语、动词合成谓语、名词合成谓语)。

传统俄语句法理论对句子类型的分析通常建立在述谓理论(предикативность)基础上。述谓核心(предикативное ядро)由主语和谓语构成,也叫述谓结构。根据述谓核心的多少区分简单句和复合句,只有一个述谓核心的是简单句,有两个或以上的述谓核心的是复合句。

基于以上分析原则和分析体系,对俄语对应单位进行语法标注,形成俄语对应单位的语法标注库,目前包括1292个俄语对应单位的标注。

4 俄语对应单位的语法分类与统计

4.1 句子和句子组成部分

汉语小句的俄语对应单位可能是句子,也可能是句子的组成部分。句子具有独立的表述功能,以句号等句末点号为收尾标点,又含独句(例(6))和句组(例(7))两类。句子的组成部分不具独立表述功能,需联合其他组成部分共同构成完整的句子,以逗号等句间点号为收尾标点,又含并列复合句组成部分,无连接词句组成部分和各种主句、从句等。

(6)^{C1} 人民群众关心的医疗、教育、养老、食品药品安全、收入分配、城市管理等方面问题较多,^{C2} 严重雾霾天气在一些地区时有发生。

^{R1}Осталось немало беспокоящих народные массы проблем в области медобслуживания, образования, обеспечения достойной старости, безопасности пищевых продуктов и медикаментов, распределения доходов, городского управления и т.д. ^{R2}По-прежнему острой остается ситуация с загрязнением окружающей среды, некоторые районы время от времени окутывал густой смог.

(7) 促进大数据、云计算、物联网广泛应用。

Последовательно продвигать вперед массовую предпринимательскую и инновационную деятельность. Стимулировать широкое распространение технологий больших данных, облачных вычислений и Интернета вещей.

图1为俄语对应单位的句子/句子组成部分统计分布,分析可知:汉语小句的俄语对应单位以句子组成部分多(967, 74.85%),句子少(325, 25.15%),前者约为后者的3倍。说明汉俄翻译中,汉语小句多译为俄语句子组成部分。

4.2 无述谓核心

无述谓核心的俄语对应单位主要存在于句子组成部分中,包括名词短语(8),动词短语(9),前置词短语(10),副词短语(11),副动词短语(12)和形动词短语(13)。

(8)^{C1} 积极的财政政策注重加力增效,/^{C2} 扩大结构性减税范围, //^{C3} 实行普遍性降费, //^{C4} 盘活财政存量资金。

^{R1}Активная финансовая политика с акцентом на наращивании интенсивности и повышении эффективности / ^{R2}позволила расширить сферы структурного сокращения налогообложения, //^{R3}перейти к тотальному снижению денежных сборов // ^{R4}и оживить неиспользованные госбюджетные резервы. (名词短语)

(9)^{C1} 加快推进产业结构优化升级,/^{C2} 实施一批技术水平高、带动能力强的重大工程。

^{R1}В связи с этим нужно ускорять оптимизацию и модернизацию структуры производств, /^{R2}реализовывать серию важных программ, имеющих высокий технический уровень и играющих мощную стимулирующую роль. (动词短语)

(10) ^{C1} 为应对持续加大的经济下行压力, / ^{C2} 我们在区间调控基础上, 实施定向调控和相机调控。

^{R1}В противовес прессингу со стороны нисходящего тренда в экономике / ^{R2}на основе определения разумных пределов нам удалось осуществить целенаправленное и своевременное регулирование. (前置词短语)

(11) ^{C1} 部分企业生产经营困难, ^{C2} 地区和行业走势分化, ^{C3} 财政收支矛盾突出, ^{C4} 金融等领域存在风险隐患。

^{R1}Часть предприятий сталкивалась с трудностями в своей производственной и хозяйственной деятельности, / ^{R2}сильно расходились тренды развития разных регионов и отраслей, / ^{R3}налицо острые противоречия между госбюджетными доходами и расходами, // ^{R4}базовые институты рынка капитала далеки от совершенства, наблюдались потенциальные риски в финансовой и других сферах. (副词短语)

(12) ^{C1} 保持经济中高速增长, ^{C2} 推动产业迈向中高端水平。

^{R1}Поддерживая умеренно высокие темпы экономического роста, ^{R2}стимулировать выход всех отраслей экономики на средний и высокий уровень развития. (副动词短语)

(13) ^{C1} 今年要实现大病保险全覆盖, / ^{C2} 政府加大投入, // ^{C3} 让更多大病患者减轻负担。

^{R1}В текущем году предстоит полностью охватить все сельское и неработающее городское население страхованием на случай серьезных заболеваний, / ^{R2}соответственно увеличатся правительственные расходы, // ^{R3}направленные на облегчение бремени еще большего числа пациентов с серьезными заболеваниями. (形动词短语)

图 2 为无述谓核心的俄语对应单位的统计分布, 无述谓核心的俄语对应单位共计 357 个, 占汉语小句俄语对应单位总数的 27.63%。分析可知: 在无述谓核心的俄语对应单位中, 以动词 (183, 51.26%) 居多, 副动词短语 (68, 19.05%)、前置词短语 (62, 17%) 和名词短语 (36, 10.08%) 居中, 形动词短语 (7, 1.96%) 和副词短语 (1, 0.28%) 较少。由此可见, 俄语中动词、副动词和前置词短语通常能够表达汉语小句的意义, 翻译中应把握它们之间的转化关系。

4. 3 单一述谓核心

单一述谓核心是指句中只包含一组主谓结构。单一述谓核心的俄语对应单位包括简单句和各种句子的组成部分。根据主谓表达手段的不同, 单一述谓核心句法单位可以是单部句: 无人称句 (14)、不定式句 (15)、确定人称句 (16), 也可以是双部句: 简单谓语 (17)、动词合成谓语 (18)、名词合成谓语 (19)。

(14) ^{C1} 深化中外人文交流, / ^{C2} 加强国际传播能力建设。

^{R1}Наряду с углублением гуманитарных обменов Китая с другими странами / ^{R2}нужно наращивать коммуникационные возможности в международных масштабах. (无人称句)

(15) ^{C1} 促进两岸文教、科技等领域交流, / ^{C2} 加强两岸基层和青年交流。

^{R1}Активизировать межбереговые обмены в таких областях, как культура и образование, наука и техника и др. / ^{R2}Расширять низовые и молодежные обмены между двумя берегами. (不定式句)

(16) ^{C1} 让我们更加紧密地团结在以习近平同志为总书记的党中央周围, / ^{C2} 凝心聚力, // ^{C3} 奋发进取, ……

^{R1} Давайте еще теснее сплотимся вокруг ЦК партии во главе с его генеральным секретарем товарищем Си Цзиньпином, / ^{R2} сосредоточим все наши силы // ^{R3} и со всей энергией устремимся вперед,..... (确定人称句)

(17) 过去一年, 我国发展面临多重困难和严峻挑战。

В прошлом году развитие Китая пережило немало трудностей и суровых вызовов. (简单谓语)

(18) 我们要持之以恒, 建设天蓝、地绿、水清的美丽中国。

Мы должны прилагать неустанные усилия к построению прекрасного Китая под голубыми небесами, с зелеными лугами и лесами и чистыми водами. (动词合成谓语)

(19) ^{C1} 取消存款利率浮动上限, / ^{C2} 推出存款保险制度, / ^{C3} 建立人民币跨境支付系统。

^{R1} Был снят верхний предел колебания процентных ставок по депозитам, / ^{R2} введена система депозитного страхования, / ^{R3} создана система трансграничных платежей в китайских юанях. (名词合成谓语)

图3是汉语小句对应的俄语句法单位可能含有多个述谓核心: 并列复合句(20)、主从复合句(21)、无连接词复合句(22)和句组(23)。分析可知: 单一述谓核心的俄语对应单位共计892个, 占汉语小句俄语对应单位总数的69.04%。分析可知: 单一述谓核心的俄语对应单位中, 以简单谓语(284, 31.84%)最多, 无人称句(185, 20.74%)、名词合成谓语(214, 23.99%)和不定式句(143, 16.03%)其次, 动词合成谓语(63, 7.06%)和确定人称句(3, 0.3%)最少。这说明, 汉语小句翻译成俄语简单句或其他句子组成成分时, 除一般动词外, 多借助无人称系词, 名词合成谓语和不定式动词。

4. 4 多述谓核心

汉语小句对应的俄语句法单位可能含有多个述谓核心: 并列复合句(20)、主从复合句(21)、无连接词复合句(22)和句组(23)。

(20) ^{C1} 完成铁路投资8000亿元以上、公路投资1.65万亿元, / ^{C2} 再开工20项重大水利工程, // ^{C3} 建设水电核电、特高压输电、智能电网、油气管网、城市轨道交通等重大项目。

^{R1} Капиталовложения в железные дороги превысят 800 млрд. юаней, а инвестиции в автодорожное строительство составят 1,65 трлн. Юаней. / ^{R2} Начато строительство 20 новых крупных гидросооружений, // ^{R3} будут реализованы крупные проекты в сфере гидроэнергетики и ядерной энергетики, электропередачи сверхвысокого напряжения, интеллектуальных электросетей, строительства сетей нефте- и газопроводов, городского рельсового транспорта и т. д. (并列复合句)

(21) ^{C1} 把适合作为地方收入的税种下划给地方, / ^{C2} 在税政管理权限方面给地方适当放权。

^{R1} Некоторые виды налогов, доходы по которым целесообразно должны были бы поступать в местные бюджеты, будут зачислены в категории местных налогов. / ^{R2} Кроме того, нужно в разумной мере увеличивать полномочия местных правительств в области налогового администрирования. (主从复合句)

(22) 合理确定增值税中央和地方分享比例。

В частности, продвигать реформу по разграничению полномочий и расходных обязательств Центра и регионов, установить для них разумные пропорции распределения поступлений от НДС. (无连接词复合句)

(23) 对行政事业性收费、政府定价或指导价经营服务性收费、政府性基金、国家职业资格, 实行目录清单管理。

Предстоит установить перечень административных сборов, перечень платежей за эксплуатационные услуги, тарифы которых назначаются или предлагаются правительством,

перечень правительственных фондов, а также список национальных профессиональных квалификационных уровней. Ускоренно формировать единую, открытую, конкурентную и упорядоченную рыночную систему в интересах преодоления регионального протекционизма. (句组)

此外,研究发现在复合句中还存在一些嵌套现象,如(24)中,整体上,R1与R2标注为并列复合句组成部分;述谓核心上,R1为表目的的主从复合句。

(24) ^{C1}我们既要让人民的物质生活更殷实, ^{C2}又要让人民的精神生活更丰富。

^{R1}Словом, нужно сделать так, чтобы материальная жизнь народа стала более зажиточной, ^{R2} а духовная – более богатой и насыщенной.

图4为汉语小句对应的俄语多述谓核心类型的统计分布。由图可知:多述谓核心类型的俄语对应单位既可存在于句子中,也可存在于句子组成部分。其中,主从复合句(23,47.92%)居多,并列复合句(8,16.67%)其次,句组(6,12.5%)和无连接词复合句(6,12.5%)为少。这说明,当汉语小句需要译为多述谓核心结构的俄语对应单位时,主要译为主从复合句。

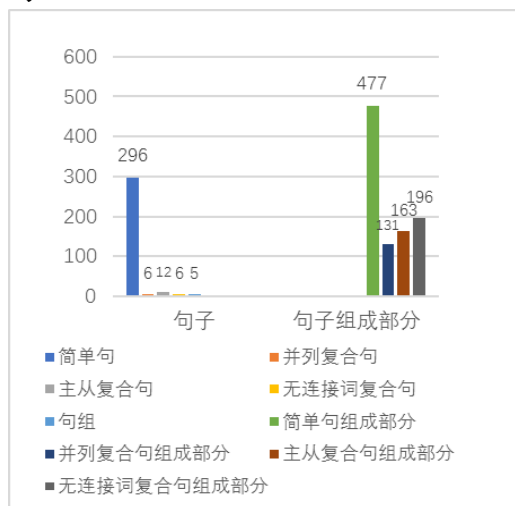


图1 句子/句子组成部分的统计分布

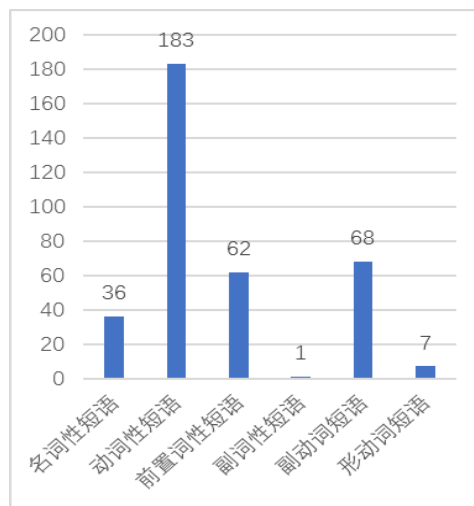


图2 无述谓核心类型的统计分布

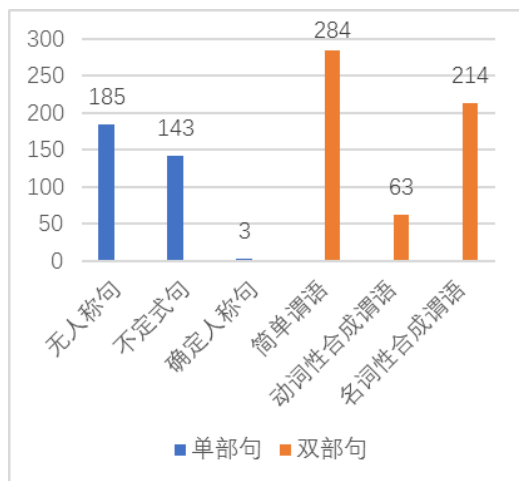


图3 单一述谓核心单位类型统计分布

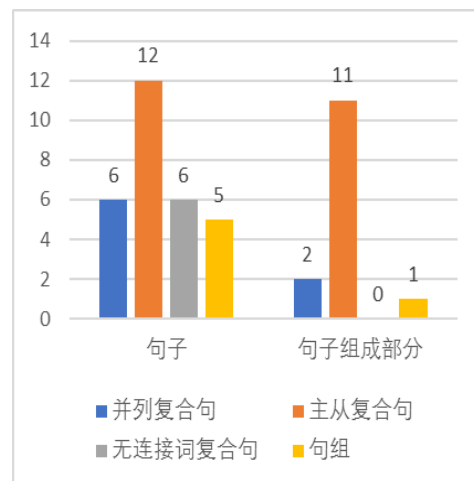


图4 多述谓核心类型的统计分布

4.5 俄语对应单位的综合统计

俄语对应单位在语法类型上呈现为9大类,在述谓核心上呈现为10大类,其分布情况见表1。分析可知:1)汉语小句译为只包含有一个述谓核心的简单句或句子组成部分时,多数译为双部句(561,43.42%),少数译为单部句(331,25.62%);其中,只有在汉语

小句译为简单句组成部分时，单部句（114，8.82%）多于双部句（79，6.11%）；

2) 汉语小句多译为单一述谓核心的俄语对应单位（892，69.04%），很大程度上多于无述谓核心（357，27.63%）和多述谓核心（43，3.33%）。可以看出，由汉语小句切分的俄语对应单位多呈现为单一述谓核心的简单句。这说明，简单句是汉俄翻译的重点。

3) 汉语小句译为简单句组成部分时，多以无述谓中心的短语（279，21.59%）表示。其中，动词短语最多（144，11.15%），副动词短语（51，3.95%）和前置词短语（50，3.87%）其次，形动词短语最少（5，0.39%）。可以看出，除在句中作同等谓语的动词短语外，通常是借助副动词、前置词、形动词等手段构成的单位。由此，用副动词、前置词、形动词等来表达汉语谓语概念，是汉俄政府文本翻译的特点。

4) 汉语小句可译为包含多个述谓核心的俄语复合句（37，2.86%）或句组（6，0.47%），其中：主从复合句多（12，0.93%），并列复合句（6，0.47%）和无连接词复合句（6，0.47%）其次，句组（5，0.39%）少；这说明，总体上汉语小句在译为俄语时，通常会含有一个或多个述谓核心。

表 1. 汉语小句对应俄语单位的述谓特征与句法类型分布表

句法类型		无述谓核心 (357, 27.63%)						单一述谓核心 (892, 69.04%)		多述谓核心 (43, 3.33%)		总计	
		名词 短语	动词 短语	前置词	副词	副动词	形动词	单部句	双部句	复合句	句组	数量	百分比
述谓核心	简单句	0	0	0	0	0	0	104	192	0	0	296	22.91%
	并列复合句	0	0	0	0	0	0	0	0	6	0	6	0.46%
	主从复合句	0	0	0	0	0	0	0	0	12	0	12	0.93%
	无连接词复合句	0	0	0	0	0	0	0	0	6	0	6	0.46%
	句组	0	0	0	0	0	0	0	0	0	5	5	0.39%
句子组成部分 (325, 25.15%)	简单句组成部分	29	144	50	0	51	5	114	79	5	0	477	36.92%
	并列复合句组成部分	3	7	5	0	8	1	21	81	5	0	131	10.14%
	主从复合句组成部分	2	29	5	0	6	0	53	67	1	0	163	12.62%
	无连接词复合句组成部分	2	3	2	1	3	1	39	142	2	1	196	15.17%
总计		36	183	62	1	68	7	331	561	37	6	1292	
		2.78%	14.16%	4.80%	0.08%	5.26%	0.55%	25.62%	43.42%	2.86%	0.47%	100%	

5 说明与总结

(1) 汉俄小句对齐标注语料库是汉英小句对齐标注语料库研究^[8]的延伸与发展。一方面，随着汉俄机器翻译的发展，目前在句子或段落层面对齐的平行语料库无法提供足够的对齐信息，所以需要基于汉俄平行语料进行进一步的对齐与分析；另一方面，未来在汉语小句

切分的基础上，试图直接建立俄英对应句法单位的对齐。

(2) 本文的分析体系仅适用于汉-俄单方向文本的研究，即汉语为源语，俄语为目的语。这是由于：汉俄小句的结构与形态有重要差异，汉译俄与俄译汉的小句转换并不相同，因此，有必要区分翻译方向，建立俄-汉小句对齐语料库，具体为：在俄汉平行语料上，切分出俄语小句，以俄语小句为标准切分汉语对应单位，形成俄汉小句对齐语料库，从而构建小句层面的双向对齐。小句（*кляуза*）不是传统俄语句法研究单位，为此，需要提出具有操作性的俄语小句定义。

(3) 本文中对俄语对应单位的语法性质的标注限于结构层面的讨论，主要是在传统俄语句法分析框架内，对句法单位的分析和讨论主要是基于结构和意义层面。下一步工作中，需要对各个句法单位在篇章功能上进行分析。

(4) 本文仅标注汉俄小句对齐语料库中俄语对应单位的语法性质，这对于指导汉语小句选择何种形式的俄语小句进行转换有一定的意义，不过，无法确定汉语小句在何种形式下转换为哪种俄语对应单位。为此，在之后的工作中，将对汉俄小句对齐语料库中的汉语小句也进行语法性质标注，以建立句法标注的对应关系。此研究除了应用在以小句为翻译单位的研究外，还将指导在整句范围内构建翻译模型。

(5) 创建汉俄小句对齐标注语料库是创建汉俄篇章结构语料库^[6-7]的基础，是篇章结构视野下的篇章-句法联合标注。小句对齐与其语法标注与篇章结构对齐将共同服务于汉俄篇章机器翻译、语言对比研究及第二语言教学。

参考文献

- [1] 王经益.面向计算机的英语关系从句汉译研究[D].北京:北京语言大学,2009.
- [2] 宋 柔,葛诗利.面向篇章机器翻译的英汉翻译单位和翻译模型研究[J].中文信息学报,2013,29(5):125-135
- [3] 陶源.人文社科学术文本俄汉平行语料库的创建与研究[J].语料库语言学.2014(01):78-93
- [4] 崔卫,张岚.俄汉翻译平行语料库及其应用研究[J].解放军外国语学院学报,2014,(01):81-87
- [5] 朴哲浩,林立娟,严文颖.基于双语新闻平行语料库的汉语新词俄语译规律研究[J].燕山大学学报(哲学社会科学版),2016,(01):138-144
- [6] YANG Y. Theoretical aspects of building a Chinese-Russian Parallel Corpus with discourse-structure annotation[A].New Russia: Traditions and Innovations language an science of language[C]. Ekaterinburg; Moscow: The Cabinet Scientist, 2016:651-660
- [7] MUKHIN M.Yu., YANG Y. Building a Chinese-Russian Parallel Discourse Structure Corpus of Official Texts[J]. Bulletin of the South Ural State University. Ser. Linguistics. 2016:23-31
- [8] FENG W., YANG Y., LI Y., REN H. Study on the English Corresponding Unit of Chinese Clause [A]. Springer: NLPCC-ICCPOL, 2016:129-140
- [9]李艳翠,冯文贺,周国栋等.基于逗号的汉语子句识别研究[J].北京大学学报:自然科学版,2013(1):7-14.
- [10]LI Y., FENG W., SUN J., KONG F., ZHOU G. 2014.Building Chinese discourse corpus with connective-driven dependency tree structure[A]. In Proceedings of EMNLP, 2014
- [11]冯文贺.汉英篇章结构平行语料库的对齐标注研究[J].中文信息学报, 2013:158-165
- [12]王磊.试论俄语“共戴句”[J].中国俄语教学, 1995:21-24

杨毅 Russia, Ekaterinburg, Ural Federal University, ul. Lenin 51, 620000 xwyang@mail.ru

冯文贺 广东省广州市广东外语外贸大学语言工程与计算实验室 510006 wenhefeng@gmail.com