上海译文出版社图书销售情况研究

高玄君

（湖南大学，湖南省、长沙市，410082）

摘要：本文使用后羿采集器，通过抓取在当当网上检索“上海译文出版社”出现的前 100 页图书信息内容，结合 Python 分析工具，从图书销售概况、读者阅读反馈两个主要方面浅析上海译文出版社纸质图书的网络售卖情况。此外，结合其他出版社的网络销售图书数据，本文把上海图书译文出版社的图书销售情况与其他大型出版社做出了简要对比，分析互联网时代纸质图书的生存之道及未来发展方向。

关键词：上海译文出版社；纸质图书；用户

中图分类号： I **文献标识码：** A

上海译文出版社是我国的老牌出版社之一，前身是上海新文艺出版社和人民文学出版社上海分社的外国文学编辑室，正式成立于 1978 年 1 月 1 日，以出版外文翻译类书籍为主营业务，是我国最大的综合性专业翻译出版社。当前，出版社主营系列包含“译文纪实”、“译文社科”、“译文电子书”等，拥有包括翁贝托• 埃科、村上春树、纳博科夫、伊恩• 麦克尤恩等在内的诸多世界著名作家的版权，出版书籍内容丰富、种类繁多。1992 年，上海译文出版社曾被国家人事部、新闻出版署命名为新闻出版系统先进单位，是中宣部表彰的全国首批优秀出版社之一。尽管作为一家拥有着较为悠久的发展历史和发展经验的大型出版社，在今天新媒体发展势头迅猛，媒体转型与融合不断推进，网络书籍普及广泛的传播语境下，上海译文出版社也要面对实体纸质书籍的营销和竞争这样的现实问题。在互联网经济日益发达的今天，除了传统的实体书店，各种网络平台也成为了许多出版社合作的选项。为了研究当下上海译文出版社书籍经营的情况，思考大型出版社当今的经营策略以及纸质图书在融合媒体时代的生存之道。本文以后羿采集器和 Python 为主要的数据挖掘和数据分析工具，从当当网这一平台抓取了上海译文出版社前 100 个详情页面的图书销售数据进行分析研究，并和其他一些出版社进行一定的对比，分析上海译文的图书出版情况以及读者反馈情况。

1. 图书销售概况分析

1.1 营销宣传的策略：基于原始词云图和统计词频的分析

在使用后羿采集器采集数据后，将未经处理的书籍标题数据利用 Python 绘制词云图，得到上海译文出版社销售书籍原始词云图（图 1），从图中我们可以观察看出，很难直接从未经处理的数据中直接分辨出出版社的图书销售类别和内容。这主要是因为当当网书籍类商品的标题，很少有只是图书名字本身的，大部分都采取了书籍名称和书籍简介相结合的标题组合方式。而且详情页在第 50 页之后，多是一些借助当当平台独立经营的个体商家，不是当当自营的内容，因此标题会带有显著的广告推销和解释自身贩卖规则的成分。观察多家出版社的销售书籍原始词云图，可以发现一样的状况。

图 1.1 上海译文出版社销售书籍原始词云图

通过分析原始词云图所呈现出的状况，可以看出，商家在表述自身售卖规则以及宣传书籍时，一些内容是经常被提及的。直观从词云图所呈现的结果来看，比较重要的一些词汇有“在线客服”、“联系”、“正版”、“全力”、“无忧”等。此外，使用 Python 对原始数据做关键词提取和词频统计，基于 TextRank 算法，词频排名前五位的分别是：“正版（1.00）”、“保障（0.68）”、“图书（0.57）”、“译文（0.54）”、“书籍（0.53）”；基于 TF-IDF 算法，排名前五位的词语分别是：“上海译文出版社(0.44)”、“正版(0.31)”、“购书(0.19)”、“无忧(0.19)”、“客服（0.18）”。对比分析这些结果我们可以看出，虽然每种算法统计的词频和词云图所呈现的结果各有侧重，但是一些共性的词汇在各种统计方式里都占据了重要的位置，最为显著的一个词汇就是“正版”，可以看出，大多数商家不论在标题中设计的宣传内容侧重哪一方面，多多少少都会提及自家书籍的正规性和合法性，是“正版书籍”，这与当下知识产权保护相关内容的推进和普及，以及大众版权意识的加强是密切相关的。此外，基于关键词词频和词云图，我们可以看出一些互联网商家在网络销售时所思考和看重的一些营销策略。

（1）提示书籍来源，增加可信度

词频统计显示的结果，不完全是形容词性和与营销行为有关的名词性的词汇，有一些与书籍本身相关的名词，同样在词频统计中占有重要的位置，如“上海译文出版社”、“译文”、“图书”、“书籍”等，这类词汇的出现，说明了商家对于书籍本身的描述性信息的重视。因为互联网上的产品纷繁复杂，读者浏览购物网站时速度往往是非常快的，而如果是不经常光顾的店铺，对其产品多少会产生一些不信任感，面对这样一种被动状态，商家只有尽可能提示一些关键信息，如书籍来源等，尤其是一些大型出版社的信息，才能使读者在有限的浏览时间内尽可能全面地形成对书籍的印象，增强信任感，增加消费的可能性。

（2）强调服务对接，争取好评数

除了在标题中说明图书的必要信息之外，很多商家也会同时在标题中注明“有问题联系客服”、“在线客服咨询”等字样。词频统计结果中，基于 TF-IDF 算法，“客服”一词出现的频率也比较高，这说明了互联网图书商家对于服务的重视。目前在网店业务体系中，商家评分是消费者重点参考内容之一。此时，对于本店服务的宣传和提示，有助于事先表明本店的服务能力和服务态度，让消费者有直观的印象感受，其次在消费者完成消费行为时，由于事先已经在标题的宣传中提示了“联系客服”的相关内容，如果消费者对产品有任何的疑问和

后续的问题，这类宣传有助于促进店家和消费者的进一步沟通，从而减少消费者随意差评的可能性。

（3）注重换位思考，优化体验感

除了以上的词汇之外，在高频词中，还可以看到一些诸如：“保障”、“无忧”等可以用来形容消费者体验感受的词汇，这些词语也是网络图书商家在宣传中喜欢使用的词语。这类词汇的显著优势在于，它们不仅是商家保证的内容，更从一定程度上反映了消费者在消费过程中重视并期望得到的感受，商家这里将这些带有主观感受色彩的词汇写进宣传广告之中，遵循了换位思考的逻辑，暗中向读者强调了在本店购物的优良体验，进一步争取希望得到良好消费体验而进行消费的顾客。

1.2 书籍销售的类型：基于处理后数据词云图和统计词频的分析

从原始数据中分析过销售策略之后，为了更清晰直观地了解到上海译文出版社本身出版和参与销售的图书类型，本文将原始数据全部导入 Word 文档，使用“替换”功能，进行广告内容的删除和书籍本身信息的提取，最终把处理好后的数据信息整理成编码为 UTF-8 的 txt 文档，使用 Python 进行云词图的绘制，得到上海译文出版社处理后销售书籍词云图（图 2）。从图中可以看出，之前的广告信息已经通过数据处理被筛选干净，这张的词云图呈现出的内容，主要是图书作家、作家国籍、图书分类以及图书名称等信息，相比于处理之前的数据，能更清楚地了解到上海译文出版社的图书经营范围和侧重点，也能更好地为上海译文出版社的在售图书进行以关键词为核心的画像。此外，通过关键词提取和词频统计的方式，基于 TextRank 算法，高频词中较为常见的词主要是有关书籍类别和体裁的词，如“小说（0.79）”、“经典（0.75）”、“文学（0.65）”、“名著（0.23）”、“故事（0.20）”；而基于 TF-IDF 算法，高频词中常见的多为一些作家的名字，其中排行前五位的作家分别是：“村上春树（0.12）”、“杜拉斯（0.07）”、“米兰• 昆德拉（0.06）”、“纳博科夫（0.04）”、“伊恩• 麦克尤恩（0.04）”，观察发现，在这次的词频统计中，TextRank 算法，虽然也可以看出一些出版社在题材上的倾向，但整体不如 TF-IDF 算法所呈现出的内容显著，基于这一算法的结果，以作家为线索，可以看出近年来上海译文出版和销售图书的一些简要情况。

图 1.2 上海译文出版社处理后销售书籍词云图

（1）选材丰富多元，时空分布广泛

在基于 TF-IDF 算法的词频统计中，词频最高的五位作家分别来自五个不同的国家——

村上春树来自日本，玛格丽特• 杜拉斯来自法国，米兰• 昆德拉来自捷克，纳博科夫来自美国，伊恩• 麦克尤恩则来自英国。不难看出，上海译文翻译的书籍并不局限在某一单一的地区或是国家，世界各地的文学精品以及经典名著都在其翻译、出版、销售的范畴之内。除了空间维度上的“分布广泛”之外，从时间这一维度来看，上海译文也拥有着广泛的涉猎和布局，从古代的莎士比亚到近现代的加缪、三岛由纪夫，再到当代的石黑一雄、村上春树，上海译文出版的图书跨越古今，涵盖了文学发展的各个时间段，内容丰富。从题材上来看，已出版的图书有文艺类、科幻类、童话类、哲学类等，题材内容也十分多元。总体来说，上海译文出版社出版社以出版外国文学作品为主要经营业务，覆盖范围广，时间跨度大，题材内容多，流行文学和经典内容并行不悖，是主推综合性外国文学的大型出版社。

（2）文学优势显著，纪实社科较弱

根据词频统计结果可以看到，词频较高的五位作家，都是小说作家，这证明纯文学类作品，尤其是小说内容，是上海译文出版社在销售时优势最为显著的一个方面。结合其官方微博“@上海译文”平时对书籍进行的宣传和推广可以看出，在纯文学之外，社科类著作、纪实文学也是出版社自身相当重视的两部分内容，然而，词频统计结果显示，这两部分的市场占有率并不理想。在前 200 个高频词中，甚至没有社科类或纪实类相关的词汇，基本都是小说作家的相关内容。产生这一现象的原因有多个，其一，虽然“译文纪实”和“译文社科”有自己独立的系列，系列中的图书也不止一本两本，但与纯文学类的其他作品的数量相比，仍有显著的差距。每个与译文合作的作家都拥有自己独立的系列，而这些作家又以小说类的纯文学作家为主，再加上之前译文历史悠久的“新经典系列”之类的作品，整体数量上，小说类相比于社科类和纪实类来说，具有压倒性的优势。其二，通过当当网、淘宝网等多个购物网站的对比观察可以发现，纪实和社科类的单本最大销量，远不如小说类的单本最大销量，从整体上来说这两类书籍有自己特定的受众群体，不如小说类整体的普适性强，因此，在销售的时候，除了官方之外的个体商家，也会相对更多地保留更为畅销的书籍内容。未来，上海译文还需要调整营销策略，在小众书籍的营销上精准定位受众群体，进行精准营销，充分发挥小众书籍的价值和影响力，使书籍的发行和销售的整体情况达到有机统一的状态。

2. 读者阅读反馈分析

使用后羿采集器对当当网前 100 页的用户评论数据进行采集，每本图书采集其首页短评，一共得到 8433 条有效数据，通过 Python 的 SnowNLP 算法进行情感分析，可以看出大部分读者对于译文出版的图书都持有正面的评价。使用 Python 绘制词云图，句型词云图和分词型词云图的呈现效果分别如图 3、图 4 所示。

图 2.1 读者评价反馈句型词云图

图 2.2 读者评价反馈分词型词云图

从句型词云图呈现的结果可以看出，读者在装帧方面，对于上海译文出版社给予了积极的评价，“装帧精美”这样的形容，出现在了词云图当中，此外，大部分读者对于书籍内容也给予了积极的肯定，肯定书籍“值得一读”、“值得一看”，还有一些读者属于“先评后看”的类型，在没有进行阅读之前就对书籍做出了评价，评论的内容中表示自己“还没看”，也有一些读者购买精装书进行收藏，表示译文出版的书籍“值得收藏”。分词型词云图呈现的结果则更为抽象，通过观察可以发现，读者更倾向于对小说类和故事型的书籍做出评价，并且村上春树这一作家，是读者容易进行阅读后反馈的热门作家，此外，博尔赫斯和麦克尤恩，也在一定程度上受到了反馈评论的读者的关注。

然而，在压倒性的好评之外，负面评价依然存在，不论在当当网这一平台上读者给出的具体评价如何，Python 的 SnowNLP 算法在进行情感分析时还是检测到了部分具有负面情感的评价，且多数“负面评价”与评价本身所表述的内容相吻合。本文摘取了部分算法识别的“负面评价”，同样对他们进行了云词图的制作，两类不同词云图的呈现效果分别如图 5、图 6 所示。

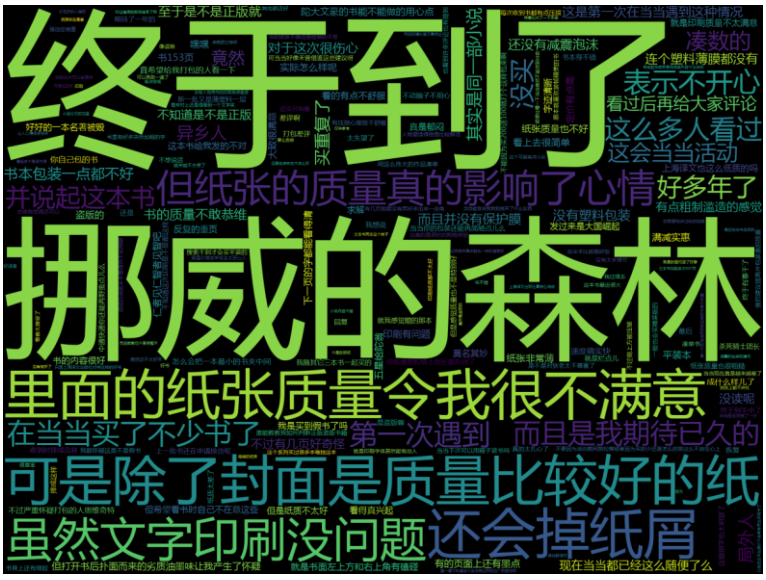

图 2.3 读者负面评价句型词云图

图 2.4 读者负面评价分词型词云图

可以看出，在负面评价方面，纸张和印刷两个大的方面，是读者不满的集中爆发点，由此许多读者会批评图书质量并因此质疑图书是否真的是正版书籍。在许多负面评价中，村上春树作品《挪威的森林》是备受争议的一本书籍，围绕它的讨论也十分丰富，此外，还有一些读者对快递和物流表示不满意。当问题发生的时候，可以看到，在物流、售后、包装、价格等方面，许多读者容易对平台本身产生不满，评论内容直接带上“当当”的名称，对平台进行批评。在印刷、翻译、内容等方面，部分读者对出版社表示了失望，认为“译文”翻译有失水准或是印刷水平略显粗糙，不够符合心理预期。

未来，译文还要面临激烈的市场竞争和较大的销售压力，在进行书籍经营的时候，有必要及时接纳读者的批评和反馈，及时改进出版过程中的失误内容，注重版权的保护，严厉打击盗版行为，以更好的内容和更高的纸质书籍质量反馈读者，并尽量扩大自己的固定读者群体。

3. 中国出版社的发展策略

基于对上海译文出版社自身的书籍数据分析，本文也采集了其他四家出版社的图书销售数据并进行简单的词云图画像，对各家出版社的主营书籍业务进行大致的对比，共同探讨当下中国出版社和纸质图书未来的发展方向。

图 3.1 上海文艺出版社：名著经典为主

图 3.2 人民文学出版社：课标教辅为主

图 3.3 新星出版社：推理小说为主

图 3.4 译林出版社：名著教辅为主

通过词云图可以看出，虽然都处在同样的出版环境之下，但各家出版社书籍出版的内容和重点还是有所区别的，通过词云图就可以有一个直观的印象。在互联网新媒体飞速发展、电子书不断扩张的今天，纸质图书和他们身后的出版社又该怎样面向未来，对于这一问题，有一些普适性的策略，对各家出版社来说，都值得思考。

3.1 明确自身定位，进行专业化经营

在纸媒和数媒共存的当下，读者和用户越来越呈现出了一种融合的趋势，即便如此，两者中也各有一部分群体是更倾向于某一类特定媒介的，也就是说纸质书依然有着一部分忠实读者，针对这一类特定群体的阅读习惯和阅读特点，各家出版社应找准定位，精准营销，避免同行之间的同类竞争、恶性竞争，彼此留下合理的生存发展空间，明确自己的出版优势和

销售优势，充分了解自己所面对的这一类读者的阅读习惯，做到有目的、有计划、有选择的出版和销售。

3.2 延伸产品链条，探索多种可能

除了传统的纸质图书的出版之外，各家出版社也可以根据自身的实际情况，有效经营自己的新媒体平台，甚至创办自己的移动应用，上海译文出版社在这方面有许多不错的尝试，新媒体方面，出版社会定期举办一些线上活动，如“朗诵<亚当夏娃日记>免费送书”活动，在书籍版权方面，译文重视纸电版权的同步发展，有文献资料显示“从 2016 年开始，纸电同步就已经成为出版界的趋势之一，电子版也由之前的可有可无变成了纸电互相促进的良方。在与版权方谈判伊始，上海译文出版社就明确提出纸电同步授权的诉求，最终如愿同时获得了电子版和纸质版两个授权。” 同时，出版社在发行和销售纸质图书的同时，也会贩卖一些与作家相关的文创周边，吸引了大量的年轻读者，延长了产业链，创新了自己的盈利模式。

3.3 重视融合营销，线上线下结合

融合营销也是未来出版社需要重视的一个发展方向，当实体书店和纸质书籍的影响力不如往日的时候，为了宣传和推广自家的书籍，尤其是新书。各类出版社可以牵头多开展一些丰富多样的线下读书活动，同时，把线上活动和线下活动有机结合起来，双线并行，共同促进书籍的宣传和推广，扩大影响力。

4. 结语

未来，媒介融合还会进一步向纵深发展，互联网经济的规模还会进一步扩大，重视融合、把握读者、精准营销，以上海译文为代表的中国出版社会在变化中不断完善、不断发展、走向未来。

参考文献

[1] 宋祎凡.翻译出版的社会历史性视野——上海译文出版社三十年图书出版概述[J].出版广角,2013,(2):66-68.

[2] 韩卫东.充分认识版权贸易的产业价值大力提高国际版权合作水平——上海译文出版社开展国际版权合作的实践和思考[J].出版广角,2011,(9):27-29.

[3] 魏蔚.小众出版的一片好天气——上海译文出版社小众出版模式浅析[J].编辑学刊,2009,(1):80-82.

[4] 施隽南.1978 年上海翻译文学出版的变迁与特色——以上海译文出版社为中心[D].上海:上海师范大学,2013.

[5] 公文,姚东敏.融合+体验:新阅读时代图书营销策略探析——以《刺杀骑士团长》为例[J].现代出版,2019,(2):38-40.

[6]汤家芳.浅议传统出版社研发 APP 的意义和困局——从上海译文出版社"译文的书"APP 说起[J].出版广角,2017,(9):12-15.

[5] 人人都是产品经理.拼多多产品深度分析：用户分析篇（上）[OL].http://www.woshipm.com/it/2657080.html/comment-page-1

A Study of Book Sales at Shanghai Translation Publishing House

Gao Xuanjun

(Hunan University, Changsha / Hunan, 410082)

Abstract: This paper uses Hou Yi collector to capture the first 100 pages of book information that appear on Dangdang.com when searching for "Shanghai Yewen Publishing House", and combines it with Python analysis tools to analyze the online sales of paper books of Shanghai Yewen Publishing House from two main aspects: book sales overview and readers' reading feedback. In addition, combining the online sales data of other publishers, this paper briefly compares the book sales of Shanghai Translation Publishing House with other large publishers, and analyzes the survival and future development direction of paper books in the Internet era.

Keywords: Shanghai Translation Publishing House; paper books; users