·临床研究·

不同话语类型影响词汇多样性和词类分布：基于普通话失语库平台的常人研究*

邓宝梅[1] 林 枫[1,2] 陈珍珍[1] 张 芹[2] 吴亚岑[3] 周 亮[4] 江钟立[1,2,5]

摘要

目的：初步构建普通话失语库常人数据集，探讨其中不同话语类型的言语产出。

方法：根据国际失语库测试标准，采集30例健康人在不同任务类型中的多媒体话语材料，采用CLAN软件进行转录、编码和校验，并对不同话语类型的词汇多样性及词类分布进行统计分析。

结果：命题故事的词汇多样性显著高于组图叙事和单图描述；三种话语类型的词类均以动词产出量最多，其次是名词；不同的话语类型具有不同的词类分布模式。

结论：话语类型能影响健康人话语的词汇多样性和词类分布，据此可为言语障碍患者在会话水平的评估和干预提供指导。

关键词 失语库；话语类型；词汇多样性；词类分布；话语评估

中图分类号：R492, R493 **文献标识码：**A **文章编号：**1001-1242(2019)-04-0410-07

Different lexical diversities and word category distributions in different discourse types: a pilot study with mandarin-control set in AphasiaBank/DENG Baomei, LIN Feng, CHEN Zhenzhen, et al.//Chinese Journal of Rehabilitation Medicine, 2019, 34(4): 410—416

Abstract

Objective: To explore the speech performance for different discourse types by establishing a control group set for mandarin in AphasiaBank.

Method: According to the established standard protocol of AphasiaBank, multi-media samples across different discourse types from thirty cognitively healthy adults were collected. Language samples were then transcribed, coded and checked by using CLAN. Lexical diversities and word category distributions across different discourse types were then analyzed.

Result: Across discourse types, the highest lexical diversity was found for story narrative compared to the other two types of sequential pictures and single picture description. All three tasks had the most output of verbs, followed by nouns. Different discourse types had different distribution patterns of word categories.

Conclusion: Discourse type influences lexical diversities and word category distributions in cognitively healthy adults. These findings make it possible to assess and treat patients with language disorders at discourse level.

Author's address The First Affiliated Hospital of Nanjing Medical University, Nanjing, 210029

Key word AphasiaBank; discourse type; lexical diversity; word category distribution; discourse assessment

DOI:10.3969/j.issn.1001-1242.2019.04.008

*基金项目：国家自然科学基金资助项目(81672255)；江苏省科技支撑计划(BE2012675)；浙江省基础公益研究计划(LGF18H170005)；江苏省高校哲学社会科学优秀创新团队建设项目(2017STD006)

1 南京医科大学第一附属医院(江苏省人民医院)，南京，210029；2 南京医科大学第三附属医院；3 湖南省人民医院；4 浙江省人民医院；5 通讯作者

作者简介：邓宝梅，女，硕士，治疗师；收稿日期：2018-04-30

话语(discourse),是人们在社会文化语境下的语言交际活动[1]。这种鲜活的交流活动所产生的语言产物处于语篇层面。以往的失语症评估,主要利用纸质量表为工具,形式上表现为独立的语音、语义、词汇和句子分析。这种脱离语境的测试任务,本质上未触及患者的日常语言交流能力。Bryant等[2]提出,在话语或语篇层面对失语症进行评估,记录和追踪言语交流恢复过程,可以从整体上评价患者的实际语言使用能力,为实现个性化言语治疗提供客观指导,更好地提高言语训练效果。国际失语库(AphasiaBank)已为此建立了标准测试方案,并提供了统计分析手段[3—5]。目前已经纳入英语、西班牙语、意大利语、日语和粤语等多种语言[4]。相比之下,汉语普通话失语库建设才刚刚起步。本研究团队的陈珍珍等[6]已初步报道前期构建的普通话命名性失语症语料分析结果。

话语的能力表现受到不同话语任务类型的影响[7—8],如看图说话、故事讲述等等。不同话语类型涉及不同的认知领域和语言要素,有着不同的言语产出[7,9]。在话语层面,言语产出是大脑复杂思维活动共同作用的结果。对不同话语类型作系统分析,可以考查各种认知要素与语言之间的复杂关系,不仅可以指导失语症评治,还有助于早期诊断阿尔茨海默病[10—14]。因此,失语库研究的要务之一,就是阐明不同任务类型下的话语特点。

本研究借助国际失语库平台,构建汉语普通话失语库常人数据集,初步探讨不同话语类型中的常态语言特征,为研究患者话语特征提供参照。

1 对象与方法

1.1 研究对象

对健康成人进行基线资料和话语材料的采集。纳入标准:①无脑卒中、脑外伤、脑肿瘤、帕金森病、阿尔茨海默病等中枢神经系统疾病;②无影响言语语言表现的外伤史、药物史;③母语为普通话;④良好的或矫正后良好的听力和视力;⑤右利手;⑥西方失语症成套测试语言功能正常,神志清楚,能配合检查;⑦参与者知情并签署知情同意书。

排除标准:①存在执行功能、注意、记忆和视空间障碍等认知功能障碍;②构音障碍;③严重的心、肺、肾等疾病或其他严重的躯体疾病;④抑郁、精神障碍;⑤不能提供知情同意。

研究方案经南京医科大学第一附属医院伦理委员会批准(批号2016-SR-007),所有参与者均已签署知情同意书。考虑今后与卒中后失语或阿尔茨海默病进行比较,因此本研究首先纳入40岁以上健康成人30例(男女各15例),年龄40—83(56.53±8.92)岁,受教育年限6—20(10.22±3.39)年,MMSE总分27—30(28.77±1.10)分,WAB的失语商为95.3—100(98.77±0.98)分。

1.2 评估方法

所有受试者在采集语料前均需接受简易精神状态检查(minimum mental state examination, MMSE)和西方失语症成套测试(western aphasia battery, WAB)。MMSE[15]包含5个方面:定向力、记忆力、注意力和计算力、回忆能力、语言能力,总分30分,>26分为正常,可反映被检者的基本认知功能,用于筛查排除认知障碍的受试者。WAB[16]是国际通用的失语症测评量表,可排除存在言语功能障碍的受试者,失语商(aphasia quotient, AQ)<93.8分可评为失语。

1.3 采集方法

所有的语料采集均在安静无干扰的室内进行,采用录像的多媒体方式记录患者在不同交谈任务中的话语材料。采集者由专门经过培训的言语治疗师担任。被检者与采集者面对面,采集者使用统一的指导语,引导患者完成测评任务。采集素材参照国际失语库标准[17],并兼顾汉文化传统制定。语料采集方法和编译技术,均已获得国际失语库采纳(中文临床指南 http://talkbank.org/manuals/Clin- CLAN-zho.pdf)。言语表现受话语任务类型的影响,其中,关于看图说话(包括组图叙事和单图描述)和命题故事的话语类型近年来受到较多关注[8,18]。本研究通过组图叙事、单图描述和命题故事这三种话语任务,探讨话语类型对言语产出的影响。组图叙事(sequential picture),即连环画,由已排序好的几张子图组成一整张带有故事情节的大图,供受试者看图讲故事。单图描述(single picture),整面大图中只有一张带有事件场景信息的图片,供受试者看图进行场景描述。命题故事(story telling),即给予受试者

图片提示后再撤除图片，让患者通过回忆讲述以往熟悉的故事内容。

组图叙事包括“打破窗户”和“拒绝雨伞”。“打破窗户”描述了一个男孩踢足球打破人家的窗户，并击倒了屋里的台灯，惊扰到了屋里的一名男子，该男子捡起球并向窗外看。“拒绝雨伞”描述了一位学生在上学时拒绝接受母亲给他的雨伞，结果在路上真的下起了雨，他只好赶回家拿伞。单图描述包括“拯救小猫”和“洪水”。“拯救小猫”描述的是一个小女孩的猫在树上，她的父亲上树试图营救猫，但被困在树上，最后是消防人员赶来救出了猫和父亲。“洪水”描述的是洪水中救援人员营救女孩的照片。组图叙事与单图描述的图片均可在失语库网站获取（https://aphasia.talkbank.org/protocol/pictures/）。命题故事包括“龟兔赛跑”和“狼来了”。“龟兔赛跑”讲述的是兔子在与乌龟赛跑中睡了一觉，结果输给了乌龟的故事。“狼来了”讲述的是放羊的孩子跟村民开玩笑，撒谎说狼来了，几次撒谎后，狼真的来了，村民也不相信他了，结果羊被狼吃掉了的故事。

1.4 数据处理

1.4.1 语料转录：参照Brian MacWhinney提供的软件教程，利用CLAN2016版把录制的多媒体语料进行转写、编码和校验，最后生成能够用于统计分析的CHAT（codes for the human analysis of transcripts）格式文件，如图1。CLAN是一个对语音材料进行标准化转写的程序，能够对语言材料进行从语义、语法到语言的多维度全方位的管理与统计分析。

1.4.2 观察指标：利用CLAN进行语料分析，可以自动生成大量的言语特征参数。本研究通过词汇多样性（由D值表示）、各词类在不同话语类型中的词频分布情况，对不同话语类型的言语产出进行对比分析。

D值是近年来提出的用于评价话语的词汇多样性的新指标[8,19—20]。D值越大，表明词汇多样性越丰富。词汇多样性（lexical diversity），反映词汇运用的广度，体现言语表达中词汇的范围或种类，是衡量言语表达质量的重要指标[21]。以往研究通常采用形次比（type/token ratio, TTR）来衡量话语的词汇多样性。TTR即类符和形符的比值。形符（token），是指话语中的所有单词，包括重复运用过的。类符

图1 经校验后的chat文件

```
File  Edit  View  Tiers  Mode  Window  Help
@Begin
@Languages:    zho
@Participants:  PAR DengBaomei004a Participant, INV DengBaomei Investigator
@ID:   zho|DengBaomei004a|PAR|57;|female|Normal|Asian|Participant|9||
@ID:   zho|DengBaomei||INV|24;|female|Normal|Asian|Investigator|17||
@Media: DengBaomei004a, video
@Date:  28-JUN-2017
@G:    broken_window
*INV:  我 会 给 你 看 一 些 图 片 . •
%mor:   pro|wo3=I v:aux|hui4=will prep|gei3=for pro|ni3=you v|kan4=look
      quant|yi1xie1=some n|tu2pian4=pictorial_matter .
%gra:   1|5|SUBJ 2|5|JCT 3|5|JCT 4|3|PREPO 5|0|ROOT 6|7|MOD 7|5|OBJ 8|5|PUNCT
*INV:  请 你 花 点 时 间 看 这 些 图 片 . •
%mor:   v|qing3=request pro|ni3=you v|hua1=spend v|dian3=dot
      n|shi2jian1=time v|kan4=look quant|zhe4xie1=these
      n|tu2pian4=pictorial_matter .
%gra:   1|6|SUBJ 2|3|SUBJ 3|6|JCT 4|5|MOD 5|3|OBJ 6|0|ROOT 7|8|MOD 8|6|OBJ
      9|6|PUNCT
*INV:  它 们 讲 了 一 个 故 事 . •
%mor:   pro|ta1-PL=it v|jiang3=speak asp|le num|yi1=one cl|ge4
      n|gu4shi4=story .
%gra:   1|2|SUBJ 2|0|ROOT 3|6|MOD 4|5|QUANT 5|6|CLASS 6|2|OBJ 7|2|PUNCT
*INV:  你 把 它 们 都 看 一 遍 . •
%mor:   pro|ni3=you prep|ba3=object_marker pro|ta1-PL=it adv|dou1=all
      v|kan4=look num|yi1=one adj|bian4=all_over .
%gra:   1|5|SUBJ 2|5|JCT 3|5|SUBJ 4|5|JCT 5|0|ROOT 6|7|QUANT 7|5|OBJ 8|5|PUNCT
*INV:  然 后 说 个 故 事 给 我 听 &啊 . •
%mor:   adv|ran2hou4=thereupon v|shuo1=say cl|ge4 n|gu4shi4=story
      prep|gei3=for pro|wo3=I v|ting1=listen .
%gra:   1|2|JCT 2|7|SUBJ 3|4|MOD 4|2|OBJ 5|7|JCT 6|5|PREPO 7|0|ROOT 8|7|PUNCT
*INV:  故 事 要 有 开 头 经 过 和 结 尾 . •
%mor:   n|gu4shi4=story v:aux|yao4=should v|you3=have v|kai1tou2=start
      n|jing1guo4=process conj|he2=and n|jie2wei3=end .
%gra:   1|3|SUBJ 2|3|AUX 3|0|ROOT 4|5|MOD 5|3|OBJ 6|7|MOD 7|3|OBJ 8|3|PUNCT
```

（type），指不反复计算的形符数，即话语中所有不同的单词。形次比的大小与文本词汇的丰富水平和多样性成正比，是词汇多样性的一种测量工具。但是，形次比极易受话语长度的影响。随着话语长度的增加，总形符和总类符也增加，但是总类符的增加速度往往没有总形符的增长速度快，到一定话语长度时，总类符趋于饱和的状态，而总形符仍会继续增加。因此，形次比极易受话语长度的影响，在不同长度的话语样本中的可比性差。而D值可以很好地解决这个问题。它主要是通过模拟形次比随着话语长度而变化的密度曲线来衡量词汇多样性的，不仅可以减弱话语长度的影响，而且充分利用了所有数据，能够涵盖更多的样本信息。D值的换算公式[19]为

$$TTR = \frac{D}{N}\left[\left(1 + 2\frac{N}{D}\right)^{\frac{1}{2}} - 1\right]$$

N表示形符数（number of tokens）。因此，D值主要与模拟曲线的斜率有关，斜率越大，即曲线越陡，表示TTR随形符数增长而下降的速度越快，D值也就越低，即词汇多样性越低。D值可以通过CLAN的voc-D统计分析程序自动生成，而且已被不少研究证实具有良好的信度和效度[9]。

1.5 统计学分析

利用CLAN 2016版进行计算机语言分析,得到词汇多样性D值、各词类的产出量,数据导出后采用SPSS 22.0软件行统计分析。计量资料以均数±标准差表示。根据自身配伍的设计方法[9],采用随机区组设计的方差分析比较三种话语类型之间D值的差异性,差异有显著性意义时用Bonferroni法对其进行进一步的两两比较。最后,采用R3.4.0进行词类频数分布的可视化分析[22]。

2 结果

2.1 三种话语类型的词汇多样性D值比较

三种话语类型的词汇多样性D值不全相同($P<0.001$)。进一步两两比较结果显示,命题故事的D值显著高于组图叙事($P<0.001$)和单图描述($P<0.001$),组图叙事与单图描述之间D值的差异无显著性意义($P>0.05$)。见表1。

表1 三种话语类型的词汇多样性D值比较 ($\bar{x}\pm s$)

	组图叙事	单图描述	命题故事
D值	62.16±3.46①	58.68±3.25①	74.22±2.87

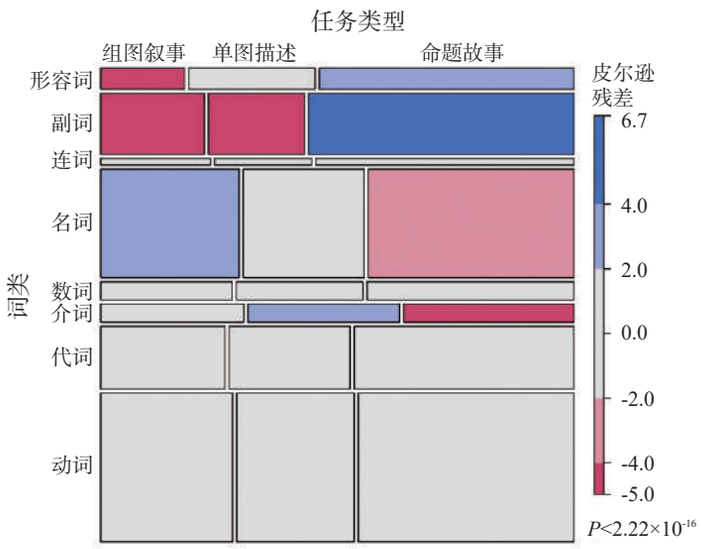
与命题故事相比:①$P<0.05$

2.2 三种话语类型的词类分布

各种词类在三种话语类型中的词频分布情况使用马赛克图展示,见图2。马赛克图是一种分类数据可视化图形,先前林枫[23]、向伟华[24]、高婧[25]等已使用该技术分析常人和患者的名词和动词语义特征分布模式。图2展示的是话语类型(3个水平)与词类(8个水平)这两个因素之间的关系。

①图2右下角$P<2.22\times10^{-16}$($P<0.05$)表示词类与话语类型有相关关系。②面积表示实际频数,各个镶嵌矩形的面积越大,则表示该词类的产出量越多。由图2可见,三种话语类型的词类均以动词产出量最多,其次是名词。这提示,动词和名词在这三种话语类型中均占据较大比例。③马赛克图中的颜色表示每个矩形的皮尔逊残差检验结果,即实际频数与期望频数之间的比较结果。蓝色矩形表示显著高于预期,红色表示显著低于预期,灰色表示无明显差异。颜色越深表示皮尔逊残差越大,颜色越浅表示皮尔逊残差越小。残差在-2.0—2.0区间之外表示概率$P<0.05$,在-4.0—4.0区间之外表示$P<0.0001$。以$P<0.05$水平来看,对三种话语类型的词类分布进行分析可知,组图叙事的名词所占频数显著高于预期,形容词和副词所占频数显著低于预期;单图描述的介词显著高于预期,副词显著偏低;命题故事的副词和形容词显著偏高,名词、介词显著偏低。综上可知,每个话语类型都有特征性的词类分布表达谱(特有的面积和颜色组合),不同的话语类型具有不同的词类分布模式,某些词类在相应的话语类型中可能会有一定的偏向性。

图2 三种话语类型的词类分布



3 讨论

对失语症患者进行评估,特别是语言在语篇语境中的实际使用表现,可以了解患者语言的实际使用能力和社会交流能力,从“生物—心理—社会”的整体层面制定个体化的治疗措施[2]。虽然当前对失语症的评估方法种类不少,但是这些评估往往侧重于音系、词法、句法等孤立的语言障碍范畴,忽略了语篇语境对语言使用的制约性[26]。同时,传统的失语症评估中,以句子为语言的最高分析层次的做法

也难以真正反映语言使用者的真实语言表现。尤其是对于轻度的失语症患者，或西方失语症成套测试诊断无失语症，但仍遗留轻度言语交流障碍的患者，传统的失语症测试往往难以检测出其语言障碍特点，也就无法为其制定相应的康复治疗方案。虽然一些评估涉及了话语层次的评价，如西方失语症成套测试中自发言语部分的场景描述，但通常情况下，这种评估是整合在整体量表里面，作为一个分项目来作分析。事实上，日常言语交流任务以及通常的言语训练中，包含了故事讲述、看图说话、自由说话等等多种话语任务类型。此外，语篇下的不同话语任务类型涉及的认知及语言要素可能不同，患者在不同话语类型中的言语产出也就会有差异[7,9]。语言功能反映大脑各种认知功能的活动状态，不同话语任务因为涉及记忆、思维、执行能力等认知要素，所以不仅可以指导失语症的评估和治疗，对早期阿尔茨海默病患者的言语认知功能的评估、训练及疗效判定也有很大帮助。阿尔茨海默病（Alzheimer's disease, AD）患者除了有记忆障碍外，语言障碍也是其早期表现之一，并随着疾病进展有不同的变化，因此语言功能是检测AD的敏感指标[10]。随着人口老龄化，AD将给家庭和社会带来沉重的负担。AD的早期防治已成为当今医学领域的研究焦点。将不同话语任务下的言语表现作为AD的早期诊断的参考，无疑有益于疾病的早期检测。话语评估目前尚未建立统一的评价指标。因此，将不同任务类型的话语分析作为临床言语障碍研究的基本层次，将在一定程度上弥补相关研究领域的空白。在这方面，国际失语库提供了一个良好的平台[3—5]，为临床言语障碍的话语研究打开了新的局面。

国际失语库是一个收集失语症患者和健康人话语的大型数据库，主要目的是为科学研究提供失语症患者和健康人在不同话语类型中的录音或录像的多媒体语料及转写材料，为全世界的失语症的治疗和研究提供大数据、多媒体、网络在线的资源共享平台[4—5,27]。它采用统一的标准测试流程，采集失语症患者和健康人的语料，利用计算机语言分析程序CLAN对语料进行转录和统计分析。本课题组借鉴国际失语库测试方案，根据文化背景对方案进行了汉化的小幅调整，目前已初步建立了汉语普通话失语症数据集，并且已经上线于国际失语库平台[6]（http://aphasia.talkbank.org/access/Mandarin/Aphasia/JiangLin.html, DOI：10.21415/T55T18）。本研究从初步建立的汉语普通话失语库常人数据集出发，对健康人在不同话语任务类型中的言语产出，以及影响话语表现的相关因素作了初期的探讨，为今后言语障碍的话语研究提供对照与参考。

本研究采用D值来衡量受检者话语的词汇多样性，可以削弱话语长度混杂因素对词汇多样性的干扰。结果显示，在三种话语类型中，命题故事的词汇多样性显著高于组图叙事（$P < 0.05$）及单图描述（$P < 0.05$）。这与Wright等[7—8,18]研究发现，在健康对照组中，命题故事的词汇多样性显著高于看图说话的结果一致。此外，Fergadiotis等[8]研究显示，组图叙事的词汇多样性显著高于单图描述。当前研究中，组图叙事的词汇多样性虽然稍高于单图描述，但差异并无显著性意义。在失语库方案中，看图说话（包括组图叙事和单图描述）是让被检者一边看图一边即兴描述故事内容；而命题故事则是先让被检者看图片，等图片拿走以后，再让被检者通过回忆复述以往熟悉的故事。命题故事的词汇多样性显著高于组图叙事及单图描述的原因，一方面，可能是因为命题故事与看图说话涉及的认知领域及语言要素不同。看图说话是让被检者一边看图一边即兴描述故事，这就要求被检者需要具备对图片故事内容的快速理解力、推理力和语言组织能力；而命题故事是复述以往熟悉的故事内容，更多的是调动被检者的记忆和复述能力。对于健康受检者来说，看图说话涉及的认知及语言要素可能更复杂，相比之下，命题故事可能显得更容易，由此产出的词汇多样性也更多。另一方面，虽然看图说话是边看图边描述，即叙述时有图片的视觉辅助，但是这样的话语类型对于健康受检者来说可能是更为限制性或封闭性的任务，反而限制了其自由表达的发挥；而命题故事虽然是在图片拿走后进行叙述，即叙述时无图片的视觉辅助，但是对健康受检者来说，可能反而是更为开放性的任务，从而使得命题故事的词汇多样性要显著高于看图说话[8]。

但是，同样的话语任务类型，对失语症患者而言，其言语产出可能又另当别论。与健康受试者不同，Fergadiotis等[8]研究显示，失语症患者的词汇多样

性在命题故事与组图叙事之间并无显著差异性。这是因为失语症患者不仅有言语障碍,还通常伴有记忆等认知问题。对失语症患者而言,图片的视觉辅助可以为其叙述提供重要的认知图解与架构。失语症患者在讲述命题故事时,因为失去图片的视觉辅助,迫使其调动更多的记忆、规划与组织等更高层次的认知加工要素[8]。根据假说,高层次的认知与低水平的词汇提取过程共享认知加工资源[28],当资源库无法承受时,会发生认知加工功能的重新分配。那么,高层次认知加工负荷的增加就必然妨碍低层次的词汇提取过程,从而导致失语症患者的词汇多样性在命题故事与组图叙事中并无显著差异。这在另一方面也提示,命题故事在失语症话语能力的鉴别诊断方面或许具有更大的潜能。在言语治疗方面,研究结果也提示我们,由于命题故事的词汇多样性显著高于组图叙事和单图描述,对于失语症患者而言,命题故事的话语类型的难度也就会高于组图叙事和单图描述,那么给患者话语训练时,就可以从组图叙事和单图描述开始着手,待改善后,再加入命题故事的训练,以逐步增加训练难度。此外,命题故事因为涉及记忆的认知成分,故可用于记忆的训练。训练可先从熟悉的故事开始,以降低记忆的难度。当患者改善后,可训练不熟知的故事,以逐步增加记忆负荷。因此,对不同话语类型进行研究分析,可以为更精准的话语评估和更个体化的言语治疗提供科学的指导。

对三种话语类型的词类分布进行分析可以发现,三种话语类型均以动词所占实际频数最多,次之是名词。这说明,动词在汉语普通话中有着重要的地位。这与动词的论元结构特点可能有关。动词不仅包含语义特征,也是语法中的关键角色,在句中起着至关重要的作用[25]。陈珍珍等[6]在普通话失语库的研究中也发现,在命题故事中,命名性失语患者的动词产出量与失语商之间呈显著正相关关系,动词的产出量更能反映命名性失语患者的严重程度。这也提示,对汉语普通话失语症患者,关注其动词的产出情况,或许对其话语功能的评价及治疗有一定的指导价值。对三种话语类型的词类分布进行比较发现,组图叙事的名词所占频数显著高于预期,这可能与组图叙事由多个子图组成,讲述者更倾向于列举图中的多种事物有关。这对临床的言语治疗也有一定指导意义,即在给患者组图叙事的训练时,给予几个关键名词的提示,或许更能引发患者对图片的叙述。单图描述的介词显著偏高,这与单图描述仅有一张图片,讲述者更倾向于从图片的各个方位或方向讲述事物间的位置关系可能有关。在临床治疗中,对于介词的词类范畴特异性损伤的患者,单图描述或许是训练的良好选择。命题故事的形容词和副词显著偏高,而在其他两种话语类型中,组图叙事的形容词和副词显著偏低,单图描述的形容词显著偏低。结合词汇多样性D值在这三种话语类型中的结果:命题故事的词汇多样性显著高于组图叙事和单图描述。这可能提示,形容词与副词的分布模式与话语类型的词汇多样性可能有关。形容词和副词均起到修饰和限制的作用,用以表示状态、程度、特征等概念,词汇丰富水平越高,则形容词和副词也可能相应地越多。在命题故事中,因为涉及熟悉的故事内容,讲述者含有以往的记忆,在讲述时更容易携带主观情感色彩,也就会运用更多的形容词和副词,其言语的词汇多样性也越高。这可能提示,副词和形容词在话语中所占构成比例的大小在一定程度上暗示着言语表达的词汇复杂程度,而且这种关系在副词中表现的更突出。这也提示我们,在临床言语训练中,提高患者对副词和形容词的运用能力,或许对改善其言语丰富度有一定帮助。

如图2所示,不同话语类型的词类分布模式具有自身的特点,词类的分布与话语类型有关。那么,在进行不同话语任务的评估时,参照健康对照组的词类分布模型和特点,可以判断失语症患者、AD患者认知受损的成分和范畴特异性损害的词类。AD患者倾向于使用更简单的语言或更少的单词,这意味着他们的词汇量减少,词汇多样性或复杂性下降,并且难以使用精确的陈述。那么,通过对比健康人在相应话语类型下的言语产出特点,可以具体了解其语言障碍的细节,甚至发展成为AD早期筛查的工具。同样,在进行言语训练时,对于不同的话语类型,可能就要考虑到患者在进行该种话语任务时需要具备什么样的认知成分、更倾向于使用什么样的词汇,从而指导治疗,为临床言语康复提供训练的线索。

本研究基于国际失语库平台,借助计算机语言分析程序CLAN的统计分析功能,对初步构建的汉

语普通话失语库常人数据集在不同话语类型中的言语产出作了初期的探索，为今后失语症的话语研究奠定了基础。然而，作为一种探索性的尝试，我们的研究还存在诸多的不足。在语料收集方面，后期的研究需要进一步扩大样本量，增强语料的完整性和代表性，纳入不同年龄层、不同教育程度的常模语料，从而深入分析年龄和教育程度等对话语表现的影响；其次，在常模构建的同时，扩充各种言语障碍类型的患者的语料，通过具体对比和分析失语症患者、AD患者等与健康人的话语表现差异，能够折射出其话语能力方面的障碍特点，帮助我们深入了解语言障碍的机制。在语料分析方面，当前研究只纳入了组图叙事、单图描述和命题故事这三种任务类型的话语分析，后期的研究可以进一步比较过程讲述、自由说话等话语类型对言语产出的影响；此外，本研究主要采用词汇多样性的D值和各词类在不同话语类型中的词频分布情况作为话语表现的评价指标，今后的研究可以纳入更多不同的话语分析方法，如主概念分析法（main concept analysis）、核心词分析法（core lexicon analysis）、连贯性分析法（coherence analysis）等，从不同的话语维度阐释语言机制，实现对语言障碍研究的整体认识。

参考文献

[1] 杨信彰.话语与语篇的研究:理论和方法[J].中国外语,2013,(4):1—18.

[2] Bryant L, Ferguson A, Spencer E. Linguistic analysis of discourse in aphasia: A review of the literature[J]. Clin Linguist Phon, 2016, 30(7):489—518.

[3] Macwhinney B, Fromm D, Forbes M, et al. AphasiaBank: methods for studying discourse[J]. Aphasiology, 2011, 25(11):1286—1307.

[4] MacWhinney B, Fromm D. AphasiaBank as BigData[J]. Semin Speech Lang, 2016, 37(1):10—22.

[5] Forbes MM, Fromm D, Macwhinney B. AphasiaBank: a resource for clinicians[J]. Semin Speech Lang, 2012, 33(3):217—222.

[6] 陈珍珍,林枫,邓宝梅,等.命名性失语的汉语普通话语料库构建[J].中国康复医学杂志,2018,33(6):669—674.

[7] Fergadiotis G, Wright HH, Capilouto GJ. Productive vocabulary across discourse types[J]. Aphasiology, 2011, 25(10):1261—1278.

[8] Fergadiotis G, Wright HH. Lexical diversity for adults with and without aphasia across discourse elicitation tasks[J]. Aphasiology, 2011, 25(11):1414—1430.

[9] Wright HH, Koutsoftas AD, Capilouto GJ, et al. Global coherence in younger and older adults: Influence of cognitive processes and discourse type[J]. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 2014, 21(2):174—196.

[10] Fraser KC, Meltzer JA, Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech[J]. J Alzheimers Dis, 2016, 49(2):407—422.

[11] Ahmed S, Haigh AM, de Jager CA, et al. Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease[J]. Brain, 2013, 136(Pt 12):3727—3737.

[12] König A, Satt A, Sorin A, et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease[J]. Alzheimers Dement (Amst), 2015, 1(1):112—124.

[13] Ferris SH, Farlow M. Language impairment in Alzheimer's disease and benefits of acetylcholinesterase inhibitors[J]. Clin Interv Aging, 2013, (8):1007—1014.

[14] Berisha V, Wang S, LaCross A, et al. Tracking discourse complexity preceding Alzheimer's disease diagnosis: a case study comparing the press conferences of Presidents Ronald Reagan and George Herbert Walker Bush[J]. J Alzheimers Dis, 2015, 45(3):959—963.

[15] 许贤豪.神经心理量表检测指南[M].北京:中国协和医科大学出版社,2007.

[16] 李胜利.言语治疗学[M].第2版.北京:华夏出版社,2014.

[17] MacWhinney B, Fromm D, Holland A, et al. AphasiaBank: Data and Methods Research methods in clinical linguistics and phonetics[M]. New York: Wiley, 2013.268—287.

[18] Wright HH, Capilouto GJ. Manipulating task instructions to change narrative discourse performance[J]. Aphasiology, 2009,23(10):1295—1308.

[19] Fergadiotis G, Wright HH, Green SB. Psychometric Evaluation of lexical diversity indices: assessing length effects[J]. J Speech Lang Hear Res, 2015, 58(3):840—852.

[20] Fergadiotis G, Wright HH, West TM. Measuring lexical diversity in narrative discourse of people with aphasia[J]. Am J Speech Lang Pathol, 2013, 22(2):S397—S408.

[21] 邓芳,郝美玲.基于“看图口语叙述”任务的泰国留学生词汇多样性发展研究[J].华文教学与研究,2017,(1):9—15.

[22] Ross I, Robert. A language and environment for statistical computing[CP/OL]. Vienna, Austria: R Foundation for Statistcal Computing, 2015. https://www.r-project.org/.

[23] 林枫.面向言语治疗的汉语词汇语义表征分析研究[D].南京:南京师范大学,2015.

[24] 向伟华,林枫,江钟立.汉语名词语义特征建模与分析[J].中国康复医学杂志,2015,30(11):1118—1124.

[25] 高婧,林枫,江钟立,等.汉语动词语义特征建模与分析[J].中国康复医学杂志,2016,(4):381—387.

[26] 赵俊海.阿尔茨海默症患者话语的系统功能语言学研究[D].重庆:西南大学,2012.

[27] Le D, Provost E M. Improving Automatic recognition of aphasic speech with aphasiaBank[J]. Interspeech, 2016:2681—2685.

[28] Rabovsky M, Alvarez CJ, Hohlfeld A, et al. Is lexical access autonomous? Evidence from combining overlapping tasks with recording event-related brain potentials[J]. Brain Research, 2008, (1222):156—165.