

# 探索语料库在民航应急服务语言中的建立和应用

毕梦丽 黄雪晴 王帅康 石杨韵雪 木丽得尔·吾拉孜汗

中国民航大学

**摘要:** 本文通过对民航应急语言研究现状及语料库在语言研究中的作用的分析,提出了民航应急语料库建设的必要性及应用路径。详细分析了语料库在民航应急语言建设中的应用、基于民航应急语言的语料库建设中的困难与限制、语料库在其他领域的成功应用与价值等诸多问题,得出关于构建基于民航应急语言服务的中英双语平行语料库的重要性及面临的挑战等相关结论。

**关键词:** 民航; 应急; 语料库

【DOI】10.12252/j.issn.2096-6288.2023.10.121

## 一、引言

语料库是指经科学取样和加工的大规模电子文本库,而其中存放的是在语言的实际使用中真实出现过的材料。语料库是语言材料,在关键时刻比如处理紧急情况时刻就会显得尤为重要。在航空领域,其可以处理例如在恐怖袭击时的各种方面的事宜,具有调整情况、稳定局势的作用等。

### (一) 背景介绍

#### 1. 民航应急语言的重要性

其他领域方面的应急语言各有所长,像是在疫情暴发期间的应急语言服务为人们保持社交距离,防范病毒传染起到了很大的作用。又或者是公共卫生事件方面总结相关日语术语进行中日翻译,从而提高翻译准确度。而在民航领域这一方面,应急语言提供了信息传递、协调国际合作、救援工作等服务,同时也能在沟通、处理家属事宜等方面发挥重要作用。

#### 2. 语料库在语言研究中的作用

一个语料库的合成或更新往往伴随着其功能的生成或改动。根据语料库的应用不同所分得的种类也不同。除开我们所说的单语、双语、多语,平行、系统语料库等,还有一些其他的语料库种类。语料库是由语言经过处理和分析过后所收集合成的语言资料合集,对于特殊情况的处理和应对有着明显的重要意义。

### (二) 目的与意义

为了能够进一步完善该语料库的系统和丰富其内容,在原有的基础上增加了新的系统和软件设置,以便后来的人可以尽量方便地使用语料库来查找相关资料。此语料库在基础功能的方面的应用不必多说,而在尽可能地完善更高级的应用功能,比如能够更好地处理紧急情况下的家属相关事宜,与国际团队沟通等。语料库目前为止还在改进的过程中,所以出现一些问题属于正常现象,我们也在为之努力。

## 二、文献综述

### (一) 民航应急语言研究现状

#### 1. 语言特点及要求

民航应急语言服务能力建设是国家应急语言能力建设的重要组成部分。而语言作为任何领域处置应急事务的重要媒介,在应急处理事件中发挥沟通作用,能有效提供应急处置手段。本文以民航应急语言建设和应用为视角,尝试建构民航应急语言语料库。民航应急语言作为行业人员重要沟通工具,具有以下特点和要求:

(1) 多样性与规范性: 民航应急语言需要适应不同文化、语境和语言环境,同时要求具备一定的规范性,以确保信息传递的准确性和标准化。

(2) 实时性: 应急情况下,信息传递需要迅速而准确,因此应急语言需要具备较高的实时性,以应对飞行中的各种紧急状况。

(3) 专业性: 民航领域有其独特的专业术语和表达方式,因此应急语言需要在专业性上达到一定标准,以确保信息传递不受误解。

总结下来,应急语言应当简洁明了、准确性高、信息传递效率高、语言标准化、紧急场景适应性、具备跨语言能力,以满足不同场合下的应急需求。

#### 2. 已有应用案例分析

标准英语通话术语的构建与颁布,有效的保障了国内外的民航飞行安全、提高了飞行效率。然而,因民航英语通讯失误造成的飞行事故远超人们的预计。例如,2017年某公司航班在香港机场起飞后至2800英尺高度开始左转,在空管管制员提示立即右转、前面有地形的情况下,管制连呼两遍后机组在3650英尺开始右转并上高度,险些导致飞机误入左边地形。这主要由于机组听错了指令,未经与管制员证实就擅自开始左转航向。由于应急语言的不规范、语言歧义、应急语言使用不当等原因导致的飞行事故,充分表明应急语言规范化和专业化在飞行事故中的重要性。

### (二) 语料库在其他领域的成功应用与价值

通过分析语料库在不同行业领域的应用，能够为民航应急语言服务的语料库建设提供一定的帮助。例如，医生和医护人员在手术室、急救室等场景中使用特定的术语和短语进行沟通，确保准确传递患者的病情和情况，以便做出适当的治疗和决策。

### （三）中英平行对齐语料库构建

中英平行对齐语料库是指使用中英两种语言撰写、相互间具有“翻译关系”的文本，这些文本在词汇、语句和语篇层面都进行了平行对齐处理。中英平行语料库的优势在于能够自动呈现两种语言的词汇、语句和语篇之间的对应关系。通过考察这些层面的对应关系，可以比较不同语言之间的异同，分析不同语言词汇之间的对应关系，探讨翻译转换规律。

#### 1、数据采集与清理

采集的数据应该涵盖不同的主题和领域，以便在翻译时能够提供更广泛的支持，在研制和语料的选取层面宜参照相关政策文件及研究成果，将民航的特性与其结合，注重应急术语的专业性表达，预设“公共卫生、灾难灾害、恐怖袭击”等类别，并且结合大量真实有效的应急语言服务事件。

在数据清理降噪阶段，通过PDF编辑器分割原文本并转换格式后使用Word软件的通配符和宏定义等功能对这部分高质量数据进行噪音处理，包括去除重复、修正错误、标准化表达等。在平行语料库中，相同的句对只应该出现一次，而不是多次，因此，需要去除重复的句对，以确保语料库的唯一性。其次，去除不匹配的句对、去除噪声和无关信息，如标点符号、拼写、语法错误等一些会影响翻译的质量的信息。对于数据源所出现的零散和不规范的现象，通过分布式计算框架如Hadoop和Python中的Pandas库来对这些能包含拼写错误、不一致的格式、噪声或缺失值的数据进行清洗和预处理。最后，需要进行人工校验，发现并纠正一些机器无法处理的问题，以确保语料库的高质量标准。

#### 2、对齐方法选择与优化

噪声处理的下一步就是对齐，在构建中英平行对齐语料库时，首先需要明确对齐的目标，例如是将源语言与目标语言进行对齐。同时，还需要确定对齐的粒度——逐词对齐、逐句对齐或逐段对齐。对齐的层面越小，技术处理的难度越大，需要根据对齐目标和对齐粒度，选择合适的工具进行对齐。对齐之后的对齐方式优化需要综合考虑多种因素，包括目标、工具、预处理、操作、验证和调整等方面。

### （四）语料库的质量评估

#### 1、准确性指标

准确性指标一般包括：词汇准确性、语法准确性、语义准确性以及语境准确性。通常采取定量和定性结合的研究方法，即：定量——基于语言资料的统计方法，可以大量而快速地收集和處理数据；采取定性的研究方法，研究应急语言服务中语言特征之间的相互作用，语言表达规律等，可以有效提高并保证研究和建设语料库的准确性。

#### 2、完整性指标

语料库的完整性指标主要是指语料库中数据内容的丰富程度和覆盖范围。具体来说，评估语料库完整性的指标可以包括以下几个方面：数据量，指语料库中数据的总量和规模，足够的语料库数据量是保证语料库完整性的基础。

### 三、语料库在民航应急语言建设中的应用

#### （一）飞机故障应急通讯

##### 1、分析应急通讯语料库的建构

语料库的建构收集、整理和分析各类语料，这些语料包括不同来源的文本、音频和视频数据，涵盖各种飞行中的紧急情况。通过使用自然语言处理技术，系统能够快速检索和分析语料库中的文本数据，使得管制人员和飞行人员能够在紧急情况下迅速获取关键信息，如飞行计划、气象条件、空域限制等。完善的语料库应用系统还应具备语音识别和转换功能，便于后续检索和分析，提高应急处理的效率和准确性。

#### （二）机组人员培训

##### 1、利用语料库改进培训材料

对收集到的数据进行深入分析，以了解民航专业人员在工作中所使用的语言、术语、交流方式等。通过自然语言处理技术，可以识别出重复出现的短语、关键词和常见的错误，这些都是培训材料中需要强调的内容。根据分析结果，可以制定更有效的培训材料，包括模拟对话、案例分析、角色扮演等。这些材料可以涵盖各种场景，如紧急情况、非正常情况、复杂天气条件等，以便让机组人员更好地了解 and 掌握相关技能。

### 四、挑战与展望

#### （一）挑战：语料库建设中的困难与限制

语料库的构建在民航领域面临着一系列困难与限制，其中主要包括数据获取难题、多语言对齐问题以及数据质量与准确性的挑战。这些问题在应急语言处理领域的研究中具有特殊性和复杂性，需要综合运用多学科知识和创新技术来解决。

##### 1、数据获取难题

在构建民航应急语言语料库时，数据获取是一项极具挑战性的任务。特别是对于涉及飞行事故、应急通讯

等敏感数据，法规和隐私问题限制了研究者获取大规模真实数据的能力。

### 2、多语言对齐问题

构建中英平行对齐语料库时，不同语言之间的表达方式和专业术语的差异性是一个显著的问题。语言之间的对应关系不仅仅包括词语，还涉及文化和专业背景的差异。这使得在对齐过程中难以保证准确性，可能导致误导性的结果。

### 3、数据质量与准确性

语料库的质量直接影响到后续应用的效果。在实际操作中，语料库可能受到错误标注、噪声数据等问题的影响，进而降低了语料库的准确性。因此，我们在构建语料库的过程中，需要采取有效的质量控制手段，包括数据清理、标注标准的统一和质量评估等，以确保语料库的可信度和可用性。

(二) 展望：结合人工智能技术的未来发展趋势和针对性语料库的构建与应用

人工智能技术的快速发展为民航应急语言处理系统的提升提供了广阔的空间。结合深度学习等新兴技术，以及构建更为专业的语料库，将成为未来民航领域语言处理的重要发展方向。

## 五、结论

### (一) 总结研究发现

#### 1、语料库在提升民航应急语言处理中的效能

民航应急语言语料库的建立对于提高飞行安全、应对紧急情况和改善跨文化交流至关重要。通过收集和整理多领域、多语种的数据，构建一个中英平行对齐的语料库，可以为应急语言处理系统提供准确、可靠的信息，从而加强飞行员、机组人员和地面控制人员之间的有效沟通和协作。

#### 2、对未来研究方向的启示

如上文所述，语料库的构建在民航领域面临着一系列困难与限制，存在一定挑战和改进空间。基于此，我们得到一些对未来研究方向的启示：

##### (1) 扩充语料库的覆盖范围

当前的语料库主要侧重于航空业务方面的文本，涉及飞行事故、应急通讯等敏感数据，获取难度大。未来可以考虑将其他相关领域的语料纳入其中，如气象数据、机械工程等，以提供更全面的数据支持。

##### (2) 改进语料库的数据质量

确保语料库的数据质量对于提高应急语言处理的准确性至关重要。未来的研究可以探索如何筛选和验证语料库中的数据，以提高数据的可靠性和有效性。如提高词汇准确性、语法准确性、语义准确性以及语境准确

性，并采取有效的质量控制手段，包括数据清理、标注标准的统一和质量评估等。

### (二) 结论与建议

#### 1、语料库在民航应急语言中实际应用的建议

在实际应用中，民航应急语言的双语语料库可以通过机器翻译、语音识别、培训与模拟演练、多语种应急指南和智能语音助手等方式发挥作用，提高紧急情况下的语言沟通效率和准确性，保障航空安全和乘客服务质量。

(1) 机器翻译与语音识别：利用语料库训练机器翻译和语音识别系统，以便在紧急情况下能够快速准确地进行双语交流。在实际应用中，这些系统可以帮助空中交通管制、机组人员和地面应急人员之间进行语言沟通，提高应急响应效率。

(2) 培训与模拟演练：将语料库中的双语内容用于培训和模拟演练，帮助相关人员熟悉并掌握应急语言的使用方法。通过模拟真实紧急情况的对话场景，可以提高人员在紧急情况下的应变能力和语言表达准确性。

(3) 多语种应急指南：基于语料库内容，可以编撰多语种的民航应急指南，为不同语种的乘客提供紧急情况下的指导和安全信息。

#### 2、语料库的重要性及未来的研究方向

综上所述，构建基于民航应急语言服务的中英双语平行语料库对于提高应急响应效率、改善应急语言服务质量以及推动智能化应急服务的发展都具有重要性。这意味着构建基于民航应急语言服务的中英双语平行语料库是一个具有潜力和重要性的研究方向。

在未来，通过结合数据收集与整合、机器翻译与对齐、领域知识与专业化应用以及跨文化交流与智能化服务等方面的研究，可以实现更好地支持民航应急语言服务的发展和应用。

## 参考文献

[1] 陈华妮, 刘梓涵, 魏珊. 国际民航组织语言能力要求下的民航应急通讯话语标准建设路径研究[D]. 四川西部文献编译研究中心: 438-431

[2] 牛原. 试析多模态语料库在跨文化交际能力培养中的应用. 文化与艺术, 2020, 09(125): 252-253.

[3] 梁砾文, 赵蓉晖. 高风险行业的“日常”应急语言规划及启示——以中国民航业为考察中心. 江汉学术, 2022, 41(3): 94-103

[4] 黄大勇, 吴土星. 民航英语通讯失误引发飞行事故分析. 中民航学院学报, 2005, 23(1): 46-48.