|  |  |
| --- | --- |
| **理事会2019年会议 2019年6月10-20日，日内瓦** | logo_C_ |
|  |  |
|  |  |
| **议项：PL 1.6** | **文件 C19/14(Rev.1) -C** |
| **2019年5月31日** |
| **原文：英文** |

|  |
| --- |
| 秘书长的报告 |
| 落实关于在同等地位上使用国际电联六种正式语文的 第154号决议（2018年，迪拜，修订版） |
| 全面实现第5号决定（2018年，迪拜，修订版）和 第154号决议（2018年，迪拜，修订版）中 关于采用备选笔译程序的条款的计划草案 |

|  |
| --- |
| 概要  理事会语文工作组（CWG-LANG）请秘书处根据第5号决定（2018年，迪拜，修订版）和第154号决议（2018年，迪拜，修订版）向理事会2019年会议提交一份**涉及未来两年或四年备选笔译、口译和字幕程序**试点项目的详细计划。  需采取的行动  请理事会将此计划草案**记录在案**，并酌情提供指导。  \_\_\_\_\_\_\_\_\_\_\_\_\_\_  参考文件  [秘书长的报告](https://www.itu.int/md/meetingdoc.asp?lang=en&parent=S19-RCLCWGLANG9-C-0002)；[第154号决议（2018年，迪拜，修订版）](https://www.itu.int/en/council/Documents/basic-texts/RES-154-C.pdf)；[第5号决定（2018年，迪拜，修订版）](https://www.itu.int/pub/S-CONF-ACTF-2018)。 |

# 1 引言

1.1 2018年全权代表大会（PP-18）批准了关于实施切实可行的措施、以节省提供口译和国际电联文件笔译费用的第5号决定。

1.2 此外，关于在同等地位上使用国际电联六种正式语文的第154号决议责成秘书长与各局主任密切协作，每年向理事会和CWG-LANG提交一份包含以下内容的报告：

• 可行的、准备由国际电联采用的**备选笔译程序**，尤其利用创新技术并了解其利弊；

• 在落实理事会所通过的**口笔译**工作的措施和原则方面所取得的进展。

1.3 CWG-LANG请秘书处在该工作组第九次会议上提交一份**今后两年或四年的备选笔译、口译和字幕程序**试点项目详细计划，其中包括工作日历、相关基准和实施该计划所需的预算拨款，以便提交理事会2019年会议（C-19）。

# 2 2018-2019年实施的项目

## 2.1 翻译程序

• 2017年，实施了一项由拉脱维亚一家语言技术公司提供的试点项目。2018年组织了试点工作，由测试和评估两个阶段组成；

• 电信标准化局（TSB）与大会和出版物部（C&P）协作，开发了一种名为“ITU-Translate”的开放源且无版税的神经机器翻译工具，并使用ITU-T技术文件对该工具进行了培训；

• 根据理事会C18/12号和C18/14号文件以及理事会2018年会议的相关决定，进行了网页翻译试验，每个目标语种有一个国际电联成员主管部门参加。

## 2.2 语音识别

2018年期间测试了一款中文语音识别产品，并准备在大会和出版物部中文科投入使用。

## 2.3 文件管理系统

前几份报告提及的由毒品和犯罪问题办公室（UNODC）在维也纳开发的文件制作系统（DCPMS）的基本版本已于2017年7月推出并在2018年初开始按照电联要求定制。

# 3 四年计划（2020-2023年）

## 3.1 目标

国际电联将提出一项为期四年的计划（2020-2023年），利用新技术（特别是人工智能和机器学习）研究并开发笔译、口译和字幕程序方面的备选程序。此计划旨在评估引入这些备选机制可能带来的效率和成本降低。这项工作将由国际电联大会和出版物部（C&P）主导，所用资金不会超出当前预算拨款的范围。

## 3.2 范围

此工作涵盖国际电联的所有语文服务：以国际电联的六种正式语文提供笔译、口译和字幕。有待调查了解的功能和工具包括但不限于：

• 规范文件的结构和标记的使用；

• 建立一个结构良好的庞大翻译记忆库；

• 质量良好的语音识别（语音-文本转换）软件；

• 机器翻译和机器学习软件；

• 文本至语音的转换软件；

• 高效且用户友好的预翻译、修订和编辑工具。

## 3.3 交付的产品

计划范围内的项目将分阶段交付，并且为便于管理起见，每年均向CWG-LANG和国际电联理事会汇报。

## 3.4 资源

国际电联需获得技能，以便能够驾驭这些新技术，可以通过向职员提供培训、参加相关会议和聘请顾问的方式实现。鉴于神经机器学习与目前可用结构的硬件要求不同，因此至关重要的是建设适当的基础设施。

## 3.5 合作伙伴关系

为在语言服务领域开展协作，国际电联希望在多个层面建立合作伙伴关系：与联合国系统、区域性组织、私营部门和单个成员国。

在联合国系统层面，国际电联已经与联合国主管大会和会议管理事务的副秘书长取得联系，以便在语文服务领域建立更加密切且正式的协作。

2019年3月，联合国与微软公司签署了一项在云计算平台Microsoft Azure的基础上开发概念验证（PoC）的协议。签署此协议旨在参与概念验证，进行内容语义分析，以确定、分类并归类联合国大会决议和全体会议报告中所包含的实体及这些实体之间关系。应当指出，联合国已经与粮农组织、劳工组织、海事组织、国际电联和儿童基金会开展了协作。

国际电联应与联合国就使用eLUNa及其他可利用AI/ML技术的翻译和语言工具进行磋商。

此外还应考虑实施联合国规范性文件和会议文件的语义互操作性框架[[1]](#footnote-1)。在此背景下，联合国于2019年初启动了两项并行举措：与微软签署了伙伴关系协议以及有关联合创意的众包挑战。

关于与区域性组织的协作，举例而言，国际电联有意与欧洲联盟开展欧洲语文网倡议（ELG）[[2]](#footnote-2)、数字创新中心[[3]](#footnote-3)和有关多语文使用的许多其他欧盟项目等议题方面的合作；例如，国际电联已作为观察员参与欧盟人工智能高级别专家组，该专家组的总体目标是支持欧洲人工智能战略的落实。

为进行语音识别和在线口译，可探索与相关私营部门实体（如微软、苹果、谷歌、亚马逊）建立伙伴关系的可能性。

可以与在提供AI/ML语言解决方案方面富有经验的成员国政府建立伙伴关系。

在具体计划和工作日历制定完成且适当费用计算结束之前，所有筹备工作可使用国际电联的节余资金进行。此外，国际电联将寻求通过捐款为部分具体项目和开发提供资金。

## 3.6 继续开发ITU-Translate和自动网页翻译

由电信标准化局管理的这一项目的范围涵盖：

• 将ITU-Translate扩展到其他部门和总秘书处。其翻译范围仅限于目前尚未翻译的文件；

• 将该工具的功能扩展到cover.pdf和.pptx格式；

• 有选择地翻译当前尚未翻译的较低层面的国际电联网页；

• 改进俄文、阿拉伯文和中文语料库的配对，以完善相关翻译；

• 通过修改算法使该工具除能够解析和处理当前的.docx和.html格式之外亦能解析并处理.pdf和.pptx格式；

• 培训该工具使其能够涵盖ITU-T技术文件以外的其它主题；

• 根据页面受欢迎程度（页面访问次数），制定确定翻译哪些网页的机制；

• 开发基于云的网页翻译功能；

• 将40 000页左右的国际电联英文网页翻译成国际电联其他五种语文；

• 向其他部门和总秘书处交付项目，包括起草用户指南和最终用户培训材料。

ITU-Translate将用于文件翻译，而外部基于云的服务将用于翻译网站上现已发布但尚未翻译的低层网页。

项目持续时间：

6个月

预算：

该项目的直接费用总额为230 000瑞士法郎，由大会和出版部（C&P）现行预算拨款出资。

## 3.7 文件管理系统

联合国秘书处正在开发一种新的文件工作流程软件 – gDoc2.0，该软件将于2019年在联合国各办事处应用，并将于2020年提供给其他客户。

## 3.8 网络翻译工具

国际电联将与日本总务省（MIC）及一家私营公司合作，测试**100多个语种的网络翻译**。

# 4 报告

所有与备选笔译、口译和字幕程序评估相关的报告均应由**备选翻译程序研究和评估小组通过CWG-LANG提交理事会**。

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. 旨在建立一个联合国范围内的机器可读文件生态系统，该系统将通过计算机可解读的数据网络，创建创新型服务，促进整个系统在信息管理方面的协作并降低成本。 [↑](#footnote-ref-1)
2. 开发首个主要的欧洲语文技术平台的项目，意味着为欧洲的语文技术解决方案打造最重要的数字市场。除公司、大学、研究中心和独立专家的商业名录外，该平台还将包含数千个数据集存储库以及数百个可在该语文网上直接使用的功能服务。 [↑](#footnote-ref-2)
3. 数字创新中心是支持企业进行数字化转型并促进其相互间合作的方式。 [↑](#footnote-ref-3)