

# 语音噪声输入对翻译性能的影响分析

田新宇, 李军辉\*

(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

**摘要:** 传统的机器翻译模型均基于无噪声环境, 即输入的数据是无噪声的。但在实际应用中, 用户输入不可避免会存在噪声, 而噪声在机器翻译过程中会直接影响其他内容的翻译, 对噪声种类及产生的影响进行统计分析对提高机器翻译的鲁棒性具有指导意义。为了模拟真实语音噪声, 本文通过人工朗读并采用讯飞语音识别系统收集 NIST 中英实验测试集的带语音噪声结果, 并对噪声文本进行统计分析, 主要包括: 1) 噪声词性分析; 2) 噪声类型分析; 3) 噪声对翻译性能的影响; 4) 噪声对其他词翻译的影响。得出的主要结论为: 名词和动词最易出现语音噪声; 同音异形类的噪声出现次数最多; 长度较小的句子在翻译时受到噪声影响的程度更加明显; 与噪声词距离更近的词翻译更易受到影响。

**关键词:** 语音识别; 噪声文本; 翻译性能; 噪声类型及影响

**中图分类号:** TP 391.2      **文献标志码:** A

## 1 引言

机器翻译 (machine translation, MT), 又称为自动翻译, 是利用计算机将一种源语言转换为目标语言的过程<sup>[1]</sup>。它是计算语言学的一个分支, 是人工智能的终极目标之一, 具有重要的科学研究价值。同时, 机器翻译又具有重要的实用价值。随着经济全球化及互联网的飞速发展, 机器翻译技术在促进政治、经济、文化交流等方面起到越来越重要的作用。机器翻译在几十年的发展过程中, 从“机械大脑”到基于规则的机器翻译、基于实例的机器翻译、基于统计的机器翻译和基于深度学习的神经网络机器翻译, 在技术上不断更新换代, 在算法上不断优化升级<sup>[2]</sup>。面对机器翻译在不同的发展状况下的优势与局限, 反思机器翻译发展面临的机遇与挑战, 提高机器翻译的鲁棒性, 在机器翻译系统快速发展的现阶段显得尤为必要。

语音识别结果作为机器翻译源语言的一种形式, 广泛应用在自然语言处理 (natural language processing, NLP) 领域中。语音识别就是让机器能够听懂人说的话, 其根本目的是研究一种具有听觉功能的机器, 使机器能直接接受人的语音输入, 理解并做出相应的反应。其基本原理是含有语音识别技术的智能体能够接收声音信号, 将信号转换成文字, 然后根据需要做记录、查询或相应的其他操作<sup>[3]</sup>。目前, 语音识别技术正趋于成熟, 同时也延伸出了诸多实用化的研究方向, 重点集中在自然话语识别与理解、实时语音识别和语音识别鲁棒性等方面<sup>[4]</sup>。

汉语语音识别的过程中, 由于朗读者自身原因以及环境因素, 不可避免地会出现一些语音噪声, 即识别不准确的词语。由于语音输入的不确定性, 识别结果中包含的噪声具有多样性, 常见的几种噪声类型有: 同音异形字、类音 (发音接近但不相同) 异形字、多字和缺字等。当语音识别结果作为翻译的输入时, 其所存在的各种语音噪声会对翻译性能产生不同程度的影响, 包括影响自身及文中其他词句的翻译。

以往不少学者对语音噪声的来源、种类, 以及噪声对某类词性词语的翻译所产生的影响进行了研究。赵晓瑞<sup>[5]</sup>提出的不规范语言翻译策略分析了各种不规范语言的来源, 指出无论是英语还是汉语都有一套统一的语言规范体系, 涉及发音、拼写、句法、修辞等各方面。不规范语言在发音上会出现错误、省略、吞音、遗漏等现象; 在句法方面表现为句子残缺不全, 不符合语法规定; 在词语方面, 会夹杂着的地域色彩, 即方言土语等。这也解释了语音识别结果中噪声词汇产生的原因。李萌<sup>[6]</sup>以信息论为依托, 以《哈利·波特》中专有名词翻译为例, 对形式噪声在翻译过程中信息传递的影响进行了分析。周美玲<sup>[7]</sup>通过对机器音译, 即人名、地名、机构名和组织名等专有名词的噪声

\* 通信作者: lijunhui@suda.edu.cn

分析，总结出基于语音的机器音译的不足的基础上，提出了两种机器音译的方法：基于字形的音译方法和机器翻译的方法。然而，常见语音噪声的类型，在语音识别结果中易出现噪声的词汇，不同噪声词汇对其他词语翻译的影响等方面还尚未被系统地统计与分析。

针对以上问题，本文基于 NIST 中英实验，以 Vaswani 等<sup>[8]</sup>提出的模型 Transformer 为基础架构，利用 NIST 中英实验中的训练集进行模型训练，采用讯飞语音识别系统获得其测试集带噪声版本。实验结合无噪声测试集和带语音噪声测试集的翻译结果进行对比，对常见语音噪声的类型，以及语音噪声词汇对其他文本翻译的影响进行了分析讨论。

## 2 数据与方法

### 2.1 数据来源

本文使用中英翻译经常使用的 NIST02 和 NIST03 两个测试集。通过人工朗读这两个测试集得到语音音频，再调用讯飞语音识别系统获得语音识别的结果，并以此作为带噪声的中文数据集。表 1 给出了两个测试集的数据统计，其中单词数和字数分别指无噪声和带语音噪声的统计值。分词均采用结巴分词<sup>1</sup>，并通过计算基于字 WER 值来衡量有噪声和无噪声数据的差异。

表 1 NIST02 和 NIST03 两个测试集的统计

Tab. 1 Statistics of the two test sets of NIST02 and NIST03

数据集	篇章数	句子数	单词数	字数	基于字 WER (%)
NIST02	95	878	18730	39988	11.24
NIST03	100	919	20041	42636	16.80

### 2.2 分析内容和方法

本文以 Transformer 为基础架构进行中英翻译模型的训练。通过对训练过程中保留的所有模型进行翻译测试，评估比较不同模型应用在开发集上（即无噪声的 NIST06 数据集）的双语互译评估（BLEU）<sup>[9]</sup>值，挑选出性能最佳的模型用于本文实验。

#### 2.2.1 Transformer 翻译模型

Transformer 架构的结构图如图 1 所示。

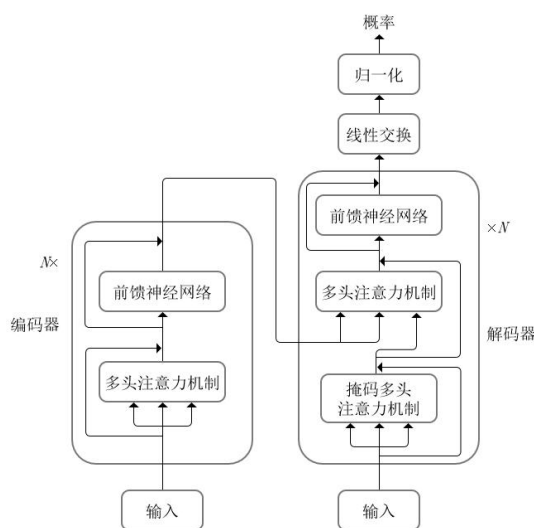


图 1 Transformer 架构

Fig. 1 Transformer architecture

Transformer 模型，用全注意力机制（attention）的结构代替了长短期记忆网络（LSTM），在以

<sup>1</sup> <https://github.com/fxsjy/jieba>

机器翻译为代表的的序列到序列任务上取得了更好的性能<sup>[8]</sup>。该模型抛弃了以往深度学习任务里面使用到的 CNN 和 RNN，广泛应用于 NLP 领域，例如机器翻译，问答系统，文本摘要和语音识别等等方向。

与 seq2seq (sequence to sequence) 模型<sup>[10]</sup>一样，Transformer 的结构也是由编码器 (encoder) 和解码器 (decoder) <sup>[11]</sup>组成。编码器由  $N = 6$  个完全相同的层堆叠而成。每一层都有两个子层。解码器同样由  $N = 6$  个完全相同的层堆叠而成。除了每个编码器层中的两个子层之外，解码器还插入第三个子层，帮助当前节点获取到当前需要关注的重点内容。子层之间均使用残差连接<sup>[12]</sup>，残差连接的方式可用式 (1) 表示：

$$h^l = h^{l-1} + f_{sl}(h^{l-1}), \quad (1)$$

其中， $h^l$  表示第  $l$  个子层的输出， $f_{sl}$  表示该层中的函数功能。

## 2.2.2 分析内容和分析方法

给定一个无噪声句子  $S = (s_1, s_2, \dots, s_n)$  共包含  $n$  个单词，以及其语音识别版本  $T = (t_1, t_2, \dots, t_m)$  共包含  $m$  个单词，为确实  $S$  中些词出现语音识别错误，本文通过以下步骤来进行原词-噪声词对的提取。

- 1) 同时遍历  $S$  和  $T$  中的每一个单词 (标点及特殊符号除外)，若当前单词  $s_i$  和  $t_j$  相同，则继续向后遍历，否则转到 2)。
- 2) 记录  $S$  和  $T$  当前噪声起始词汇的下标  $p$  和  $q$ ，并找到  $p$  和  $q$  位置之后距离两噪声起始词汇最近的相同的单词，记录其下标分别为  $x$  和  $y$ ，即为当前噪声结束词汇的下标。转到 3)。
- 3) 在  $S$  中提取出  $O = (s_p, \dots, s_x)$  的单词组成当前噪声的原词组，在  $T$  中提取出  $N = (s_q, \dots, s_y)$  的单词组成当前噪声词组，将二者结合为原词-噪声词对  $(O, N)$ 。转到 1)。

例如原句为：“以 公务员 所 共同 遵循 的 行为规范、**至理名言** 为 书法作品 创作 内容，倡导 公务员 廉洁奉公 之风。” 噪声句为：“以 公务员 所 共同 遵循 的 行为规范、**治理名言** 为 书法作品 创作 内容，倡导 公务员 廉洁奉公 **指望**。” 该句所提取出的原词-噪声词对为：“(“至理名言”，“治理名言”)和(“之风”，“指望”)。

本文首先分析哪些词容易引起语音噪声，然后再从多个角度分析语音噪声输入对机器翻译性能的影响。分析的内容包括噪声词性分析、噪声类型分析、噪声对翻译性能的影响分析，以及噪声对其他词翻译的影响分析等。

**噪声词性分析。**我们想知道哪些词容易在语音识别中产生错误。为此，我们分析每类词性的单词出现语音识别错误的比例。

**噪声类型分析。**常见的语音识别错误包括替换、缺词或多词等。有别与英文等西方语言，中文语音识别存在着大量同音异形字或者类音异形字。为此，我们人工对比原句和语音识别结果，将噪声类型具分为同音异形字、人名、类音异形字和多字少字等类别，并统计它们的比例。

**噪声对翻译性能的影响。**对比和分析无噪声和带噪声测试集的翻译性能。

**噪声对其他词翻译的影响。**当一个句子存在语音噪声时，分析当前句子的其他哪些词容易受噪声的影响。为了分析每处噪声对翻译的影响，我们要确保每个句子只有一处噪声。为此，如果某个句子出现多处噪声，针对每一处噪声生成一个句子，以保证该句中只保留一处噪声。例如上述例子中的原句和噪声句含有两处噪声，因此生成两个句子，分别为：“以 公务员 所 共同 遵循 的 行为规范、**治理名言** 为 书法作品 创作 内容，倡导 公务员 廉洁奉公 之风。”和“以 公务员 所 共同 遵循 的 行为规范、**至理名言** 为 书法作品 创作 内容，倡导 公务员 廉洁奉公 **指望**。”记  $S$  和  $T$  分别表示无噪声句子和只有一处噪声的句子，即句子  $S$  只有且只有一个单词，记为  $s_i$  在语音识别中出现错误。为了判断  $S$  中其他某个单词，如  $s_j$  ( $i \neq j$ ) 的翻译是否受噪声影响，我们分别获取  $S$  和  $T$  的译文，然后通过双语词对齐工具 fast\_align<sup>2</sup>，获取  $s_j$  在两个句子译文中的翻译结果，如果两个译文相同，即认为  $s_j$  的翻译不受  $s_i$  发生语音识别错误的影响，反之受其影响。具体地，根据与发生语音识别错误词  $s_i$  的平面或结构距离，我们分析哪些词更容易受语音识别错误的影响。以

<sup>2</sup> [https://github.com/chagge/fast\\_align](https://github.com/chagge/fast_align)

下以  $s_i$  为例，解释平面距离和句法距离。

- 平面距离：即与  $s_i$  的表面距离。如果当前词为  $s_j$ ，其与  $s_i$  的距离为  $|i - j|$ 。
- 结构距离：即与  $s_i$  在句法依存上的距离。从  $s_i$  出发，沿父结点或子结点方向能到达  $s_j$  过程中经历的边数。为了得到句子  $S$  的句法树，本文使用依存句法分析工具 DDParse<sup>3</sup>。

### 3 实验

本研究对 NIST 中英实验的 NIST02 和 NIST03 数据集中中文语料的语音识别结果进行噪声分析，结合带噪声中文语料的翻译结果来统计噪声词汇的词性、类型及其他翻译受影响词的占比。本文采用训练集语料共包括 2 百万句子对，来源于 LDC2002E18、LDC2003E07、LDC2003E14、LDC2004T08、LDC2002T01、LDC2004T07、LDC2005T06、LDC2005T10、LDC2009T02、LDC2009T15 和 LDC2010T03 等多个数据集。开发集为 NIST06。中文端使用结巴分词，英文使用 Moses 脚本对句子进行 tokenization，以及小写化。中英两端分别做子词化处理，不共享两端词表，字节对编码 (BPE)<sup>[16]</sup> 操作数均设置为 32000。

#### 3.1 实验设置

本文实验的基准模型使用开源 OpenNMT<sup>[15]</sup>实现的 Transformer<sup>4</sup>。训练时，中英模型相同设置的参数主要有：编码器与解码器的层数均为 6 层，多头注意力机制设置 8 个头，批处理大小设置为 4,096，词向量、编码器和解码器的隐藏层维度均为 512 维，前馈神经网络的维度设置为 2,048 维，dropout<sup>[17]</sup> 设置为 0.1。使用 Glorot 方法初始化模型参数，其他参数均使用默认配置。翻译模型在一块 GTX 1080Ti 显卡上训练。在网络训练过程中，采用 Adam 算法进行参数更新，其参数 beta1 为 0.9，beta2 为 0.998，训练过程中每隔 5000 步保存一次模型。在翻译测试集句子时，柱状搜索设置为 5。

#### 3.2 结果与分析

##### 3.2.1 噪声词性分析

图 2 统计了在人工朗读或语音识别过程中易出现噪声的词汇的词性占比，总噪声词数为 4810。

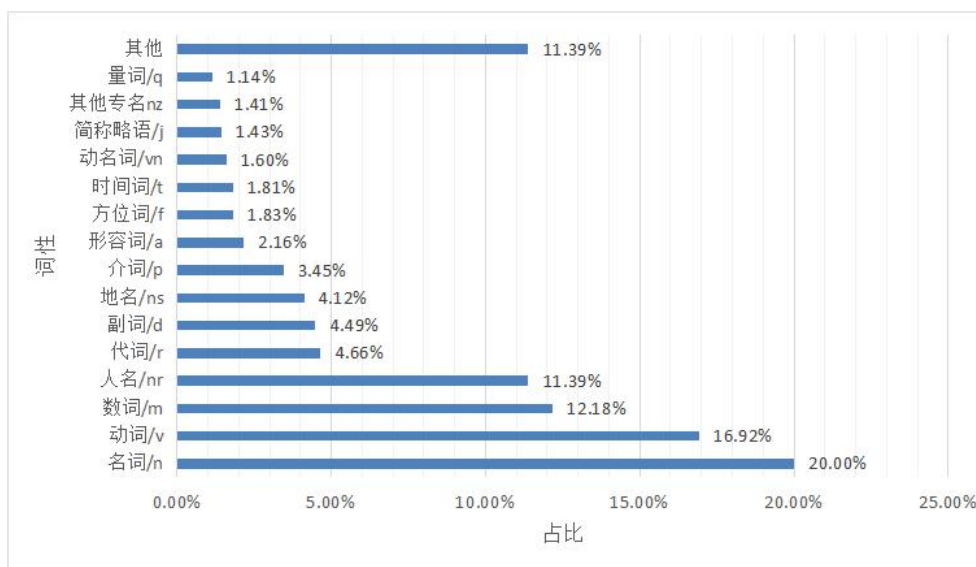


图 2 不同词性的噪声词汇占比

Fig. 2 Proportion of noise vocabulary of different parts of speech

从图 2 可以看出，实词更容易出现语音噪声，其中名词、动词、数词以及人名词性占比较大。图 2 按占比大小的顺序统计了 15 种词性，剩下更多占比较小的词性合计占比为 11.39%。这样的占

<sup>3</sup> <https://github.com/baidu/DDParser>

<sup>4</sup> <https://github.com/OpenNMT/OpenNMT-py>

比分布也比较符合我们的日常生活，实词占据了日常交流语言中的大部分。

表 2 在图 2 的基础上统计了每种词性的噪声词汇数占该词性词汇出现总数的比例，目的在于进一步分析哪些词性的词汇更易出现噪声。为了使结果更具统计意义，表中仅记录了出现总数在 1000 以上的词性。中文语料的总词数为 38771，其中包含 4810 个噪声词，占比 12.41%。从表 2 可以看出人名出现噪声的词数占总词数的比例最高，达到了 45.67%，且远高于占比第二高的代词（17.91%），这说明人名词性在朗读或语音识别的过程中最容易出现噪声。此外，数词、副词、动词、名词等占比也相对较高。

表 2 不同词性的噪声词汇占该词性总词数的比例

Tab. 2 Proportion of noise vocabulary of different parts of speech in the total number of it

词性	噪声词数	总词数	占比
人名/nr	548	1200	45.67%
代词/r	224	1251	17.91%
数词/m	586	3498	16.75%
副词/d	216	1913	11.29%
动词/v	814	7254	11.22%
名词/n	962	8722	11.03%
地名/ns	198	2435	8.13%
介词/p	166	2060	8.06%
形容词/a	77	1015	7.59%
其他	1019	9423	10.81%
<b>合计</b>	<b>4810</b>	<b>38771</b>	<b>12.41%</b>

### 3.2.2 噪声类型分析

图 3 统计了不同种类的噪声在所有噪声中的占比，总噪声数为 2806。

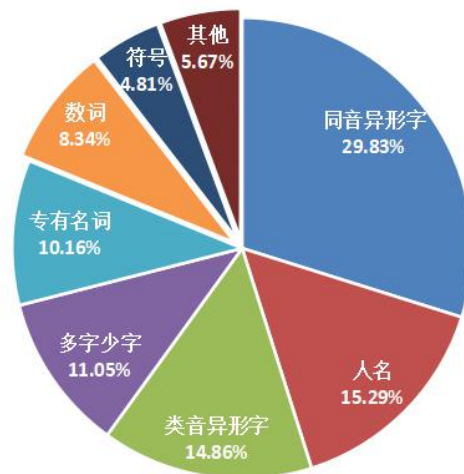


图 3 不同类型的噪声占比

Fig. 3 Proportion of different types of noise

从图 3 可以看出，主要的噪声类型包括同音异形字、人名、类音异形字和多字少字等，其中同音异形字的噪声占比最高，且远高于占比第二的人名类型的噪声。同音异形字、人名以及专有名词

的噪声类型反映了语音识别软件的不完全准确的现象，如果词汇不是很常见或者朗读音调有所偏差，那么语音识别系统较容易识别为同音异形字；而遇到人名或其他专有名词时，由于这些名词中的汉字组合无规律，因此也较易出现识别错误的情况。类音异形字和多字少字等噪声类型反映了人工朗读的不准确性，即可能受地域口音或环境因素的影响。

表 3 该出了一个原句及其语音识别结果举例。在该例中“阿姆 丹说”属于人名类型的噪声，“红星 月 会”属于专有名词类型的噪声，“进东”属于同音异形字类型的噪声，“建林”属于专有名词类型的噪声，“制定”属于同音异形字类型的噪声。

表 3 原句与语音识别结果举例

Tab. 3 An example of an input and its ASR output

原句（已分词）	哈姆丹 说 ， 负责 运作 此事 的 阿联酋 红新月会 在 与 联合国 近东 巴勒斯坦 难民 救济 和 工程处 协调 后 ， 已 就 修复 杰宁 难民营 内 被 毁坏 的 建筑物 和 公用设施 制订 了 具体 计划 。
语音识别结果（已分词）	阿姆 丹说 ， 负责 运作 此事 的 阿联酋 红星 月 会 在 与 联合国 进东 巴勒斯坦 难民 救济 和 工程处 协调 后 ， 已 就 修复 建林 难民营 内 被 毁坏 的 建筑物 和 公用设施 制定 了 具体 计划 。

### 3.2.3 噪声对翻译性能的影响

表 4 给出了带语音噪声与无噪声 NIST 中英测试集的翻译性能。从表 4 可以看出，测试集语音识别的结果的 BLEU 值比标准语料低了 10 个点左右。

表 4 带噪声与不带噪声 NIST 中英测试集的翻译性能

Tab. 4 BLEU values of NIST ZH-EN with and without noise

语料类型	NIST02	NIST03
带噪声	35.52	32.06
无噪声	46.13	41.66

图 4 展示了不同长度带噪声和无噪声句子的翻译性能，句子长度划分区间为(0, 10]、(10, 20]、(20, 30]、(30, 40]、(40, +∞)。

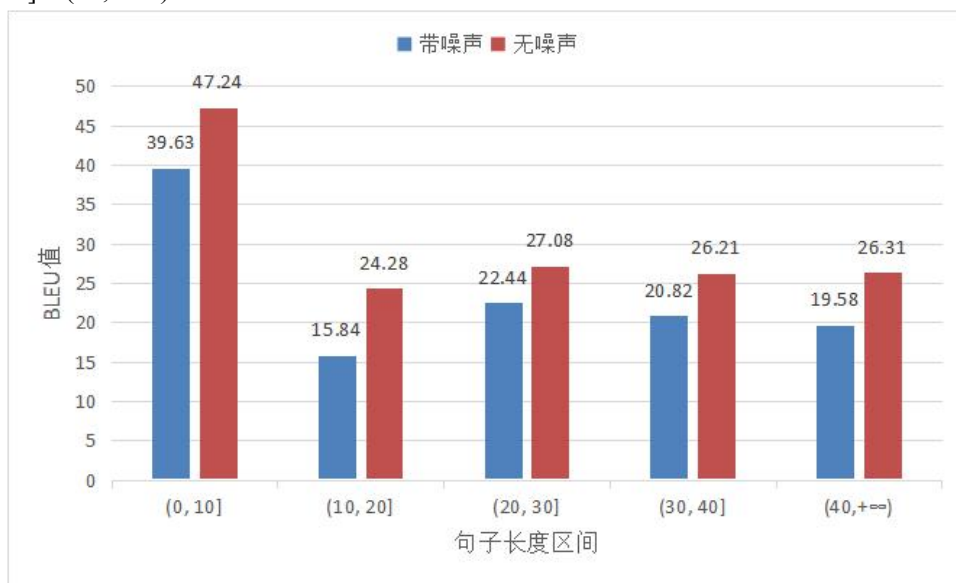


图 4 不同长度带噪声和无噪声句子的翻译性能

Fig. 4 BLEU values of sentences of different lengths with and without noise

从图 4 可以看出，长度在 20 以下的句子在翻译时受到噪声影响的程度更加明显，BLEU 值比无噪声句子的翻译低 7~9 个点；而长度大于 20 的句子在翻译时受到噪声影响的程度略有降低，BLEU 值比无噪声句子的翻译低 4~7 个点。

### 3.2.4 噪声对其他词翻译的影响

统计翻译受噪声词汇影响的正常词汇的平面与结构距离区间均为 1~5。表 5 给出了一个原句及其语音识别结果的翻译举例。在该例中，由于“伤愈”一词的识别错误，对与其平面距离为 1 的词“复出”，平面距离为 2 的词“最早”、平面距离为 3 的词“乔丹”的翻译均产生了影响；对与其结构距离为 1 的词“复出”的翻译产生了影响。

表 5 原句与语音识别结果翻译举例

Tab. 5 An example of translation of an input and its ASR output

原句及翻译（已分词）	乔丹 最早 周日 伤愈 复出。 Jordan to come back from his recovery of injury as early as Sunday .
语音识别结果及翻译（已分词）	乔丹 最早 周日 商誉 复出。 Jordan 's first trade fair breaks out .

图 5 统计了出现噪声的情况下，句中其他正常词汇翻译受影响的占比，即噪声词汇影响其他某个距离词汇翻译的次数占噪声词汇出现的总次数的百分比，总噪声数为 2806。从图 5 可以看出，当按照平面距离划分时，距离噪声词 5 以内的词翻译较容易受影响，且随着平面距离的增加，受影响词的占比逐渐降低。此外，平面距离 5 以上的词受影响占比平均为 7.90%，明显低于平面距离 5 以内的词。当按照结构距离划分时，与噪声词结构距离为 1 的词受影响最多，占比达到了 23.60%，与噪声词结构距离 2 以上的词受影响次数较少，只有极少部分词性的噪声词影响的其他词的结构距离会达到 5。此外，对比两折线趋势可以得出结论：1）平面距离、结构距离与翻译受噪声词汇影响的正常词占比均为负相关；2）随着距离的增大，结构距离对翻译产生影响降低程度比平面距离更加明显；3）噪声词汇从结构距离角度对其他词的影响总体低于平面距离。

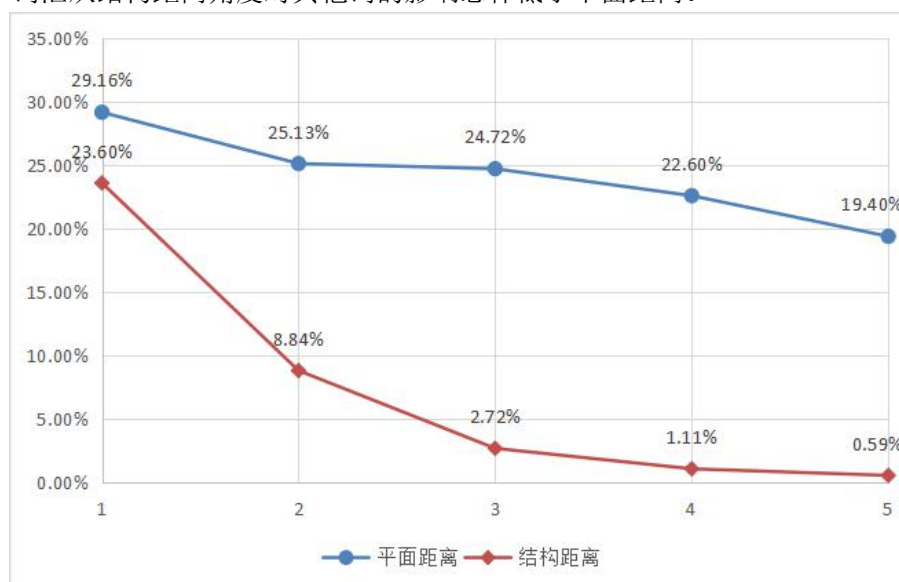


图 5 翻译受影响的正常词汇占比

Fig. 5 Proportion of the translation of normal words affected by noise

## 4 总结

本文综合利用 NIST 中英实验语料库的中英文数据集，通过讯飞语音识别系统和机器翻译，对



生成的无噪声和带噪声中英文语料进行比对分析，系统地统计了常见语音噪声的词性、语音噪声的类型、噪声对翻译性能的影响以及哪些范围的词汇翻译易受语音噪声的影响等。研究表明：

- 1) 名词和动词等实词更容易在语音识别过程中形成语音噪声；
- 2) 同音异形字的噪声出现次数最多，该类噪声反映了语音识别系统对同音异形字的识别结果不够准确。此外也存在较多人工朗读的问题造成的类音异形字和多字少字的语音噪声；
- 3) 长度较小的句子在翻译时受到噪声的影响程度更加明显；
- 4) 无论与噪声词汇在已分词的句子中的平面距离，还是与噪声词汇依存关系相关的结构距离，其他词翻译受影响的占比基本都和距离成负相关。

综上所述，不同词性、类型的语音噪声对其他词汇翻译的影响多出现在与其距离较近或依存关系较近的位置。基于此，在对中文语音识别结果进行翻译的过程中，可重点关注那些易被噪声词汇影响的词的翻译，从而进行翻译纠错，提高机器翻译的鲁棒性。

## 参考文献：

- [1] 李亚超,熊德意,张民.神经机器翻译综述[J].计算机学报,2018,41(12):100-121.
- [2] 高璐璐,赵雯.机器翻译研究综述[J].中国外语,2020,17(06):97-103.
- [3] 于俊婷,刘伍颖,易绵竹,等.国内语音识别研究综述[J].计算机光盘软件与应用,2014(10):76-78.
- [4] 马志欣,王宏,李鑫.语音识别技术综述[J].昌吉学院学报,2006,000(003):93-97.
- [5] 赵晓瑞.不规范语言翻译策略的理论透析:语域的视角[J].石家庄学院学报,2006(01):106-109.
- [6] 李萌.从信息论角度看形式噪声对专有名词翻译的影响——以《哈利·波特》为例[J].学园,2017(26).
- [7] 周美玲.英汉人名音译方法的研究与实现[D].苏州大学,2009.
- [8] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Neural Information Processing Systems. Los Angeles: NIPS, 2017: 5998-6008.
- [9] BALABANOVIC M, SHOHAM Y. Fab: content-based, collaborative recommendation [J]. Communications of the ACM, 1997, 40(3): 66-72.
- [10] Ilya Sutskever, Oriol Vinyals, and Quoc VV Le. Sequence to sequence learning with neural networks. In Advances in Neural Information Processing Systems, pages 3104–3112, 2014.
- [11] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. CoRR, abs/1406.1078, 2014.
- [12] HEK, ZHANGX, RENS, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2016.
- [13] 王步康,王红玲,袁晓虹,等.基于依存句法分析的中文语义角色标注[J].中文信息学报,2010.
- [14] Dozat T, Manning C D. Deep Biaffine Attention for Neural Dependency Parsing[J]. 2016.
- [15] KLEIN G, KIM Y, DENG Y T, et al. OpenNMT: Open-Source Toolkit for Neural Machine Translation[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. [S. I.]: Association for Computational Linguistics, 2017: 67–72.
- [16] Sennrich R, Haddow B, Birch A. Neural Machine Translation of Rare Words with Subword Units[J]. Computer Science, 2015.
- [17] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.



# Analysis of the influence of speech noise input on translation performance

TIAN Xinyu, LI Junhui\*

(School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

**Abstract:** Traditional machine translation models are based on a noise-free environment, that is, the input data is noise-free. However, in practical applications, user input will inevitably have noise, and noise will directly affect the translation of other content during the machine translation process. Statistical analysis of the type of noise and its impact is instructive to improve the robustness of machine translation. In order to simulate real speech noise, this paper collects the results with speech noise from the NIST Chinese-English experimental test set through manual reading and adopts the iFLYTEK speech recognition system, and conducts statistical analysis on the noise text, mainly including: 1) part of speech analysis of noise; 2) analysis of the types of noise; 3) the influence of noise on translation performance; 4) the influence of noise on the translation of other words. The main conclusions drawn are: nouns and verbs are most prone to speech noise; homophone noises appear the most frequently; sentences with smaller lengths are more affected by noise during translation; words that are closer to the noise words are more likely to be affected by noise during translation.

**Keywords:** speech recognition; noisy text; translation performance; types and effects of noise