



MiraiTranslate  
みらい翻訳

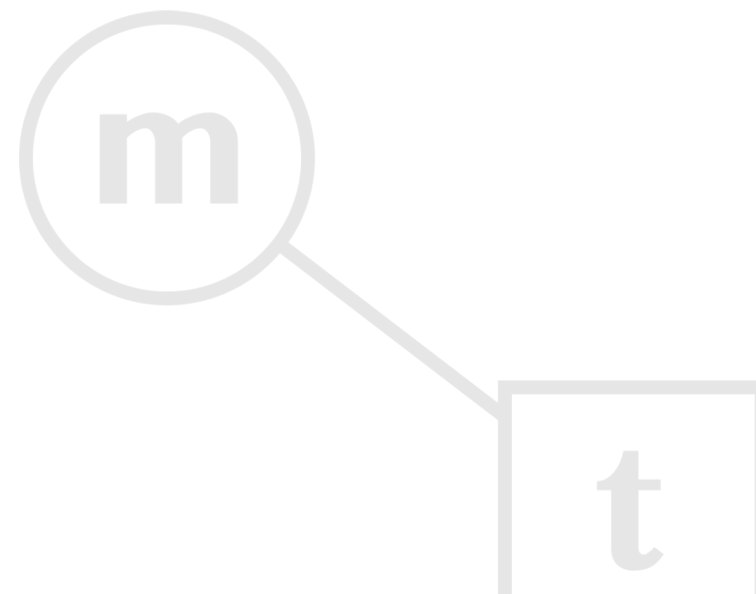
# 翻译窍门

株式会社みらい翻訳  
Mirai Translate, Inc.

〒150-0002 東京都渋谷区渋谷 2-22-3 渋谷東口ビル2F

TEL: 03-6434-1340 FAX: 03-6434-1590

<https://miraitranslate.com>



※本资料的内容可能随产品更新而有所差异

## 0. [简介](#)

### 1. [神经网络机器翻译的特点](#)

### 2. [如何让机器翻译更准确](#)

### 3. [文本翻译与文件翻译](#)

### 4. [用户词典](#)

1. 什么是用户词典?
2. 用户词典 使用窍门 (初级篇)
3. 用户词典 使用窍门 (高级篇)

### 5. [翻译记忆库](#)

1. 什么是翻译记忆库?
2. 翻译记忆库 使用窍门

### 6. [参考资料](#)

- 翻译流程

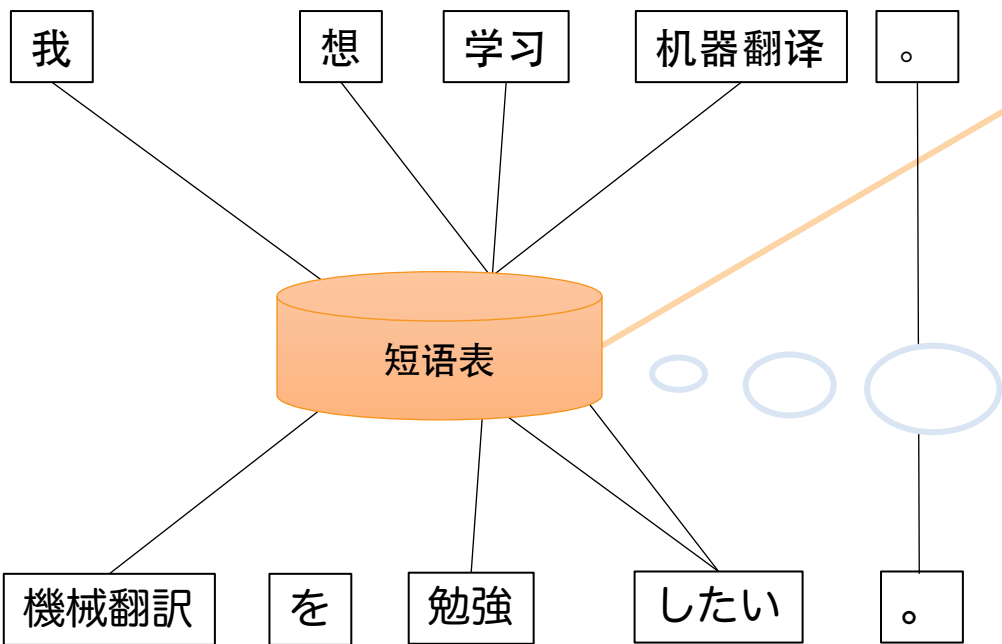
# 0. 简介 (迄今为止的机器翻译概述)



到2010年代前期为止，说到“机器翻译”的话，一般是指基于词组的统计型机器翻译(Phrase-Based Statistical Machine Translation, PBSMT)，通过单词串单位的重组来生成译文的该技术曾是主流。由于是以单词串为单位剪切粘贴而成的句子，所以整体上看存在着不流畅的问题。

## 基于传统技术 (PBSMT) 的SMT机器翻译

注意：为了便于解释，我们做了部分简化



语料库:是指进行数据库化处理的大规模的语言资料。用于机器翻译的对译语料库是将各种语言的语料库一一对应起来的。

日语	汉语
ホテルに泊まります。 水を少し下さい。 :	下榻酒店。 请给我一点水。 :

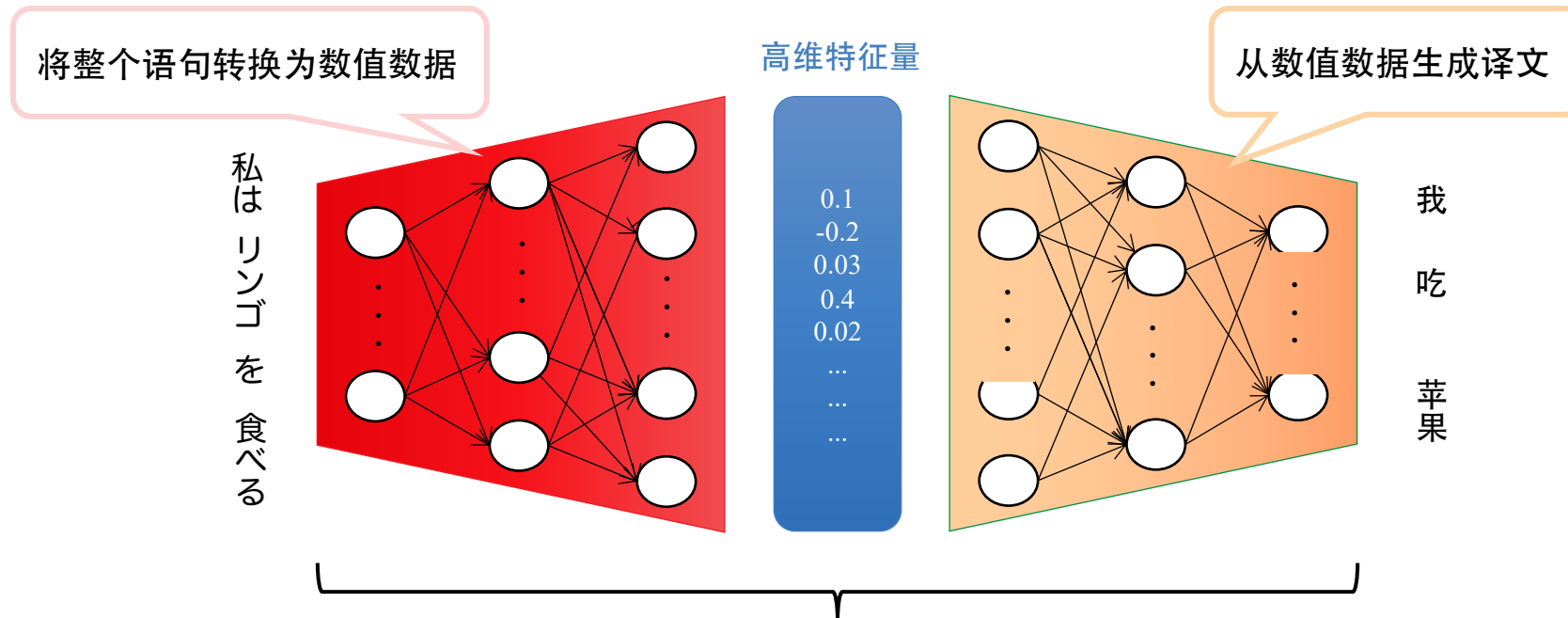
通过学习对译语料库，计算出对应关系概率最高的词组其在统计学上的概率来重新排列语序。如例：

	「想」的日语翻译	与「我」并用时对应的概率
1	したい	80%
2	思います	●%
3	考えます	▲%

“我想...”与「...したい」，通过统计学上的关联概率进行判断。

# 1. 神经网络机器翻译的特点

Mirai Translator采用神经网络机器翻译(Neural Machine Translation, NMT)。NMT不是以词组为单位，而是将整个语句转换为数值数据，然后根据该数值数据生成新的译文。与PBSMT相比，它能够提供更流畅自然的翻译结果。



它是最新的人工智能翻译引擎，它将原文替换为数值数据并生成翻译结果

# 1. 神经网络机器翻译的特点 翻译精度提高的实例



与传统的PBSMT技术相比，NMT大大提高了翻译的准确性和流畅性。

原文	我が国の産業化の歴史を長期的に見てみると、経済発展や与件の変化に伴って、異なる時代で「生きる道」は変化してきている。
----	--



PBSMT 翻译结果	我国的工业化的历史,长期来看,经济发展和条件的变化,不同的时代,“活路”发生了变化。
---------------	--

不准确不自然

Mirai NMT引擎 翻译结果	从长期来看我国工业化的历史，随着经济发展和条件的变化，“生活方式”在不同时代发生了变化。
---------------------	--

准确自然

# 1. 神经网络机器翻译的特点 需要注意的特征

翻译精度高是NMT的一大特征，与传统的技术相比，可以生成更流畅自然的文章。而另一方面，我们希望您理解NMT也存在以下需要注意的特征。

## 需要注意的特征

- ✓ 我们还不清楚它为什么能更流畅自然。  
→ 同样，翻译错误出现的时候没有明确的修复方法。
- ✓ 翻译结果不一定完全覆盖原文  
→ 可能会出现漏译、增译、过度生成等翻译问题。
- ✓ 对原文的细微差异比较敏感，同一个单词的译文容易出现不统一的情况，而且可能会根据标点是否存在而发生很大变化（不稳定性）。
- ✓ 很难完全控制和匹配词典。  
→ 使用用户词典时请仔细阅读 4. 用户词典



# 1. 神经网络机器翻译的特点 不稳定性



NMT对原文的细微差异比较敏感，同一个单词的译文容易出现不统一的情况，而且可能会根据标点点的存在与否而发生很大变化。

- 有/无标点

原文	翻译结果
有一笔技术咨询的费用急着需要付款	技術相談の料金が必要になりました
有一笔技术咨询的费用急着需要付款。	技術相談の料金が急に必要になりました。

- 不统一

原文	翻译结果
振替輸送を行います。	进行代替运送。
振り替え輸送を行います。	进行调换运输。



## 2. 如何让机器翻译更准确



机器翻译不能像我们人类一样理解上下文，因此尽量输入简单易懂的原文较能准确地翻译。比如，在将日语翻译成中文时，可以通过避免日语中常见的模糊表达，内容省略，间接措辞等来提高机器翻译的准确性。

具体而言，有以下几点：

1. **简单易懂的语序**
  - 修饰语和被修饰语尽量靠近，不要分得太开
2. **尽量避免“省略”**
  - 避免使用支离破碎的口语措辞，选择适当的书面表达
  - 注意数字的缩写
  - 不省略主语或宾语
3. **尽量避免委婉歧义** ※原文是日语
  - 明确形容的对象
  - 用汉字代替容易产生歧义的日语假名
4. **尽量避免晦涩难懂的表达**
  - 避免成语或谚语
  - 避免非常用表达
5. **避免使用非简体字** ※原文是中文简体字
  - 使用繁体字和二简字容易导致翻译错误（原文是繁体字则相反）

## 2. 如何让机器翻译更准确 具体例子



### 1. 简单易懂的语序

- 修饰语和被修饰语尽量靠近，不要分得太开

	原文	翻译结果
修改前	弊社では、本年度から昔の大きな非常にコストのかかる機械の改善に努めています。	在敝公司，从本年度开始，一直在努力改善以前非常昂贵的机器。
Good 修改后	弊社では、本年度から非常にコストのかかる大きな昔の機械の改善に努めています。	在敝公司，从本年度开始努力改善成本非常高的旧机器。

## 2. 如何让机器翻译更准确 具体例子

### 2. 尽量避免“省略”

- 避免使用支离破碎的口语措辞，选择适当的书面表达

	原文	翻译结果
修改前	そのお客様にはなるはやで電話をかけてください。	请给那位客人打电话。
Good 修改后	そのお客様にはなるべく早く電話をかけてください。	请尽快给那个客人打电话。

- 注意数字的缩写

	原文	翻译结果
修改前	2~30人	2~30人
Good 修改后	20~30人	20~30人

- 不要省略主语或宾语

	原文	翻译结果
修改前	明後日の14時半にお願いします。	后天14:30。
Good 修改后	明後日の14時半に打ち合わせをお願いします。	请在后天14:30碰头。

## 2. 如何让机器翻译更准确 具体例子



### 3. 尽量避免委婉歧义

- 不要使用有歧义的句子。

	原文	翻译结果
修改前	要死了猎人的狗。	猟師の犬が死にそうです。
Good 修改后	猎人的狗被咬死了。	猟師の犬がかみ殺されました。
Good 修改后	猎人被狗咬死了。	猟師が犬に噛まれて死にました。

- 用汉字代替容易产生歧义的日语假名

	原文	翻译结果
修改前	あめが降るといけないので傘をもってってください。	因为不能下糖，所以请带着伞。
Good 修改后	雨が降るといけないので傘をもってってください。	下雨就不好了，请带上伞。

## 2. 如何让机器翻译更准确 具体例子



### 4. 尽量避免晦涩难懂的表达

- 避免成语或谚语

	原文	翻译结果
修改前	彼女に何度も連絡したが梨の礫だ。	我跟她联系了好几次，都是梨子。
Good 修改后	彼女に何度も連絡したが返事がない。	我跟她联系了好几次，都没有回音。

- 避免高度概念化的表达

	原文	翻译结果
修改前	请开文明车做文明人。	文明車に乗って文明人になってください。
Good 修改后	请遵守交通规则，文明驾驶。	交通ルールを守って、マナーを守って運転してください。

## 2. 如何让机器翻译更准确 具体例子



### 5. 避免使用非简体字 ※原文是中文简体字

#### • 避免使用繁体字

	原文	翻译结果
修改前	昨天風很大，吹得人睜不開眼睛。	昨日はとても大きく,人から目を奪われた。
Good 修改后	昨天风很大，吹得人睁不开眼睛。	昨日は風が強くて目が開けられませんでした。

#### • 避免使用二简字

	原文	翻译结果
修改前	请问步厅在哪里。	すみません、どこですか。
Good 修改后	请问餐厅在哪里。	すみません、レストランはどこですか？

我们提供“文本翻译”和“文件翻译”两种方式。根据需求来选择使用会很方便，但是各有其需要注意的地方。

- 在执行翻译之前
  - 请确认**语言方向**是否正确。
  - 使用翻译设置来统一文体，这样的话翻译结果会更容易阅读（统一文体功能目前仅适用于翻译结果为日语时）。默认状态时翻译结果的日语文体为不统一。
- 共通的注意事项
  - 翻译是一句一句地进行，**不会考虑上下文**（不能推测句子前后的意思）。因此，**无论是逐句输入还是整段输入，翻译结果都不会改变。**
- 文本翻译的注意事项
  - 因为换行符被视为句子分隔符，所以**如果在句子中间添加换行符以调整句子的外观（调整布局）的话，翻译结果的准确性可能会降低。**  
(特别是从PDF复制文本来进行翻译时，或是在电子邮件文本中插入换行符时，请特别注意该问题)
- 文件翻译的注意事项
  - 在PDF文件的翻译中，某些文件需要通过**OCR识别字符**，如果无法正确识别，翻译结果的准确性很可能会降低。如果您持有转换为PDF之前的源文件，则将其进行文件翻译会更准确。
  - 与文本翻译一样，换行符被视为句子分隔符。特别是对PDF文件进行翻译时，识别过程中可能会插入不需要的换行符。

### 3. 文本翻译与文件翻译

首先 ① 设定语言方向，然后 ② 请确认“翻译设置”是否合适。（“翻译设置”目前仅适用于翻译结果为日语时）。

#### 文本翻译



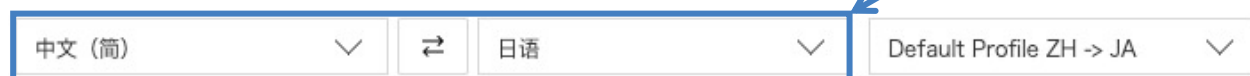
① 设定语言方向



② 确认“翻译设置”



#### 文件翻译



② 确认“翻译设置”



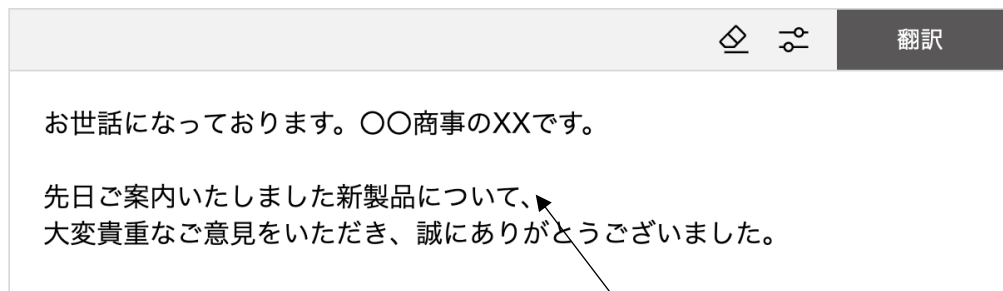


### 3. 文本翻译与文件翻译



文本翻译中包含换行符的示例如下所示。  
因为换行被视为句子结尾，所以建议删除句子中间不必要的换行符之后再翻译。

示例:为了让邮件更容易阅读而换行



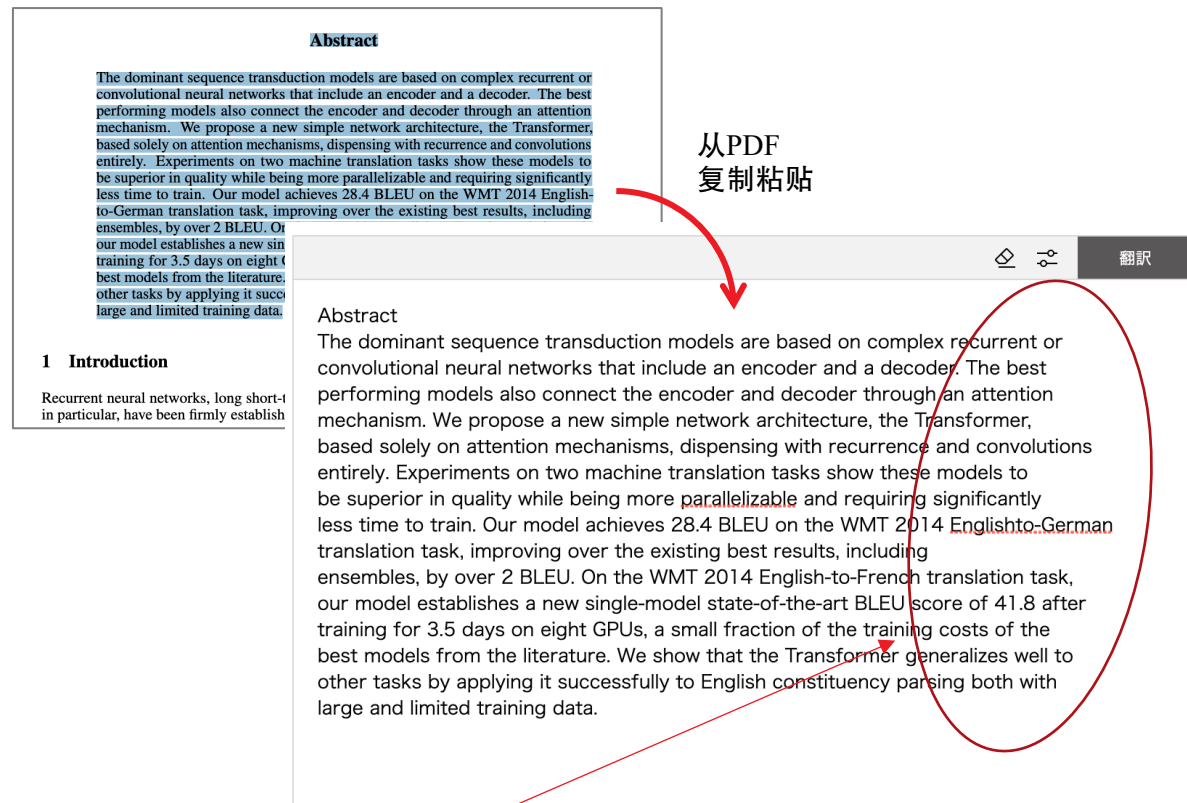
如果被换行符分成这样的，会逐行翻译

Good

先日ご案内いたしました新製品について、大変貴重なご意見をいただき、誠にありがとうございました。

删除换行符，还原成一句话，就能使翻译更准确。

示例:从PDF复制粘贴时，原始文本插入了用于调整布局的换行符



从PDF  
复制粘贴

检查每行的结尾、语句中间是否有换行。如果有换行符，删除换行符能使翻译更准确。

## 4-1. 什么是用户词典?



用户词典：是指可以按照用户的设定，翻译专有名词和专业术语的功能。  
针对在使用Mirai Translator时没能正确翻译的单词，通过把您公司内部的用语集等创建成用户词典，从而提高翻译的精度。

示例: 「サービス開発部に所属している窪田と申します。」

未使用用户词典的话:

我是服务开发部的洼田。



人名“洼田”翻译错了.....  
部门名称“服务开发部”想翻译成“SDD”.....

使用用户词典的话:

我是隶属于 SDD 的 窪田。



“SDD”、“窪田”都按照我们的要求那样翻译准确了!

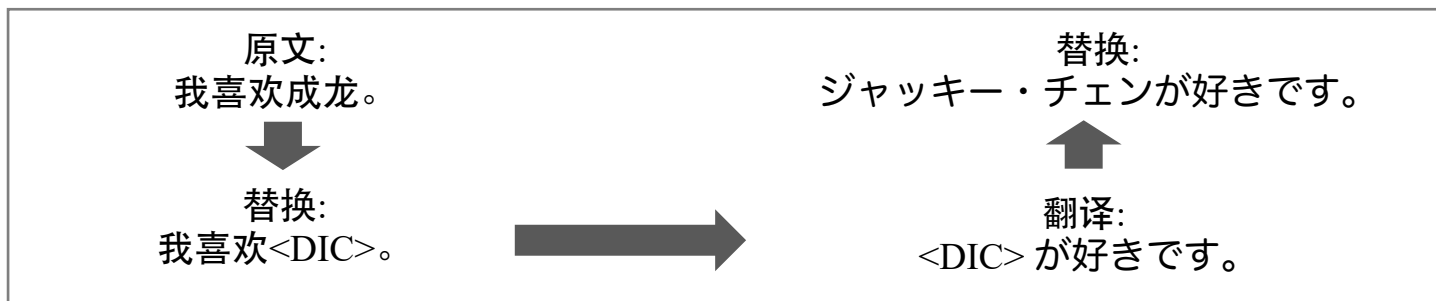
※使用用户词典时，除了登录的单词以外的部分的翻译也有可能发生变化。

## 4-1. 什么是用户词典?

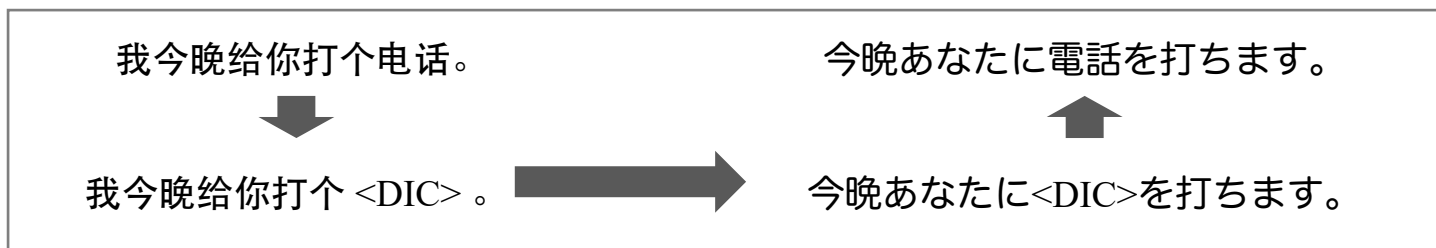
当使用用户词典功能时，匹配的部分文本会被转换为特殊标记，这可能导致翻译质量降低。为了避免这种情况，我们建议 (1) 尽量避免创建非专有名词 (2) 依次添加未能准确翻译的词语。

### • 用户词典功能的机制和注意事项

- 用户词典功能是通过原文中匹配部分暂时地替换为特殊标记(分类字段)后进行翻译，翻译后将分类字段直接替换为词典中登录的译词来实现的。(下图为创建了“成龙→ジャッキー・チェン”词典的示例)



- 因此，当匹配词典时，是在缺少部分原文含义的情况下翻译，有时可能会发生以下误译(下图为创建了“电话→電話”词典的示例)



→尽量不要创建非专有名词非常重要

※在我们的机器翻译引擎中，<DIC>被识别为专有名词。不把“电话→電話”加入词典的话，可以正确翻译为「今夜あなたに電話をかけます。」，强行加入的话反而降低翻译质量。

为了有效地使用用户词典，有以下几点需要留意。

- **理解匹配的优先顺序**
  - 文章中先出现的单词优先匹配
  - 字数多的单词优先匹配
  - 排在前面的单词优先匹配
- **尽量避免添加专有名词以外的单词(普通名词、抽象名词等)**
  - 避免添加普通名词
  - 避免添加抽象名词
- **理解会有无法正常反映的情况**
  - 单词的某一部分无法正常反映
  - 从前往后按顺序优先匹配

## 4-2. 用户词典 使用窍门 (初级篇)

### 1. 理解匹配的优先顺序

- 基本上是按照指定的顺序优先匹配，但是有些情形下顺序靠后的单词会优先匹配。匹配的优先顺序如下所述。
  - i. 在句子中首先出现的
  - ii. 字数更多的
  - iii. 指定顺序靠前的

词典1

No.	原文	翻译结果
1	東京	东京

词典2

No.	原文	翻译结果
1	東京	东京
2	東京タワー	东京塔
3	タワー大展望台	塔观景台

※按照词典1、词典2的顺序指定的情况

東京タワー大展望台は東京都港区にあります。



词典2\_2 大展望台は 词典1\_1 都港区にあります。



东京塔大展望台位于东京都港区。

①因为词典1中收录了「東京」，句中的「東京タワー大展望台」的「東京」本来应该比词典2的「东京」优先进行匹配，但是按照规则 ii.词典2中「東京タワー」的字数更多，所以会更优先匹配词典2。

②此时，比「東京タワー」字数更多的「タワー大展望台」并不会被优先匹配。这是因为按照规则 i. 在句子中首先出现的「東京タワー」优先匹配。

③关于句子中「東京都港区」的「東京」，按照规则 iii. 指定顺序靠前的词典1的「東京」被匹配。

## 4-2. 用户词典 使用窍门 (初级篇)

### 2. 尽量避免添加专有名词以外的单词(普通名词、抽象名词等)

- 避免添加普通名词

普通名词的 苹果 ⇔ リンゴ

	原文	翻译结果
Good	这个苹果好甜啊。	このリンゴは甘いですね。
No Good	什么牌子的手机比较好，苹果怎么样。	どのメーカーの携帯がいいですか?リンゴはどうですか?

※手机品牌的“苹果”，日语是「アップル」

- 避免添加抽象名词

既是名词也是动词的 骄傲 ⇔ 誇り/おごる

	原文	翻译结果
Good	灿烂的古代文化遗产是我们的骄傲。	輝かしい古文化遺産は我々の誇りである。
No Good	他的学习成绩很好，但从来不骄傲。	彼は成績はとてもよいが、これまで誇りしたことがない。

※骄傲”的动词用法，日语是「おごる」

## 4-2. 用户词典 使用窍门 (初级篇)



### 3. 理解会有无法正常反映的情况

- 只添加单词的某一部分无法正常反映

	原文	翻译结果
添加前	ガーラ湯沢駅	GALA汤泽站

- 在这里，想要把“GALA”全部改成部分小写字母，但是即使把 ガーラ ⇔ Gala 添加进词典，也无法反映在翻译结果中。(翻译系统内部认为“ガーラ湯沢”是一个单词。即使将该单词的某一部分「ガーラ」添加进词典，机器翻译也会解释成别的单词。如果想进一步了解翻译系统的结构，请查阅[6. 参考资料翻译的流程](#)。)
- 这种情况，将 ガーラ湯沢 ⇔ Gala汤泽 添加进词典，就容易正确地反映出来。

	原文	翻译结果
Good 添加后	ガーラ湯沢駅	Gala汤泽车站

如上例，即使添加到了词典中也无法反映时，如果改成更长一些的单词，就容易正确地翻译出相应的结果。

## 4-2. 用户词典 使用窍门 (初级篇)



### 3. 理解会有无法正常反映的情况

- 从前往后按顺序优先匹配

想把「都庁舎 ⇔ 都庁大楼」添加到用户词典

	原文	翻译结果
添加前	東京都庁舎	東京都官署办公大楼

此时，假如 東京都 ⇔ 东京都 也收录在词典中，用户词典功能会优先匹配最先出现的单词，所以对于「東京都庁舎」会反映出“东京都”，而不会反映出“都庁大楼”。

	原文	翻译结果
Good 添加后	東京都庁舎	東京都庁大楼

在这种情况下，应该把 東京都庁舎 ⇔ 東京都庁大楼 添加到用户词典，这样就能够正确翻译了。



使用用户词典时，需要注意以下几点。

### 1. 创建用户词典之前请确认以下几点

- **确认词性**

- 删除非名词。一般名词容易因上下文而变化，也需要尽量删除。

- **确认字数**

- 字数少的单词，与其他字数多的单词部分一致，从而被其他单词优先匹配的可能性较高，应删除。

- **确认重复**

- 同一个单词不能重复创建，应删除。

### 2. 创建之后试着检查翻译结果是否有变化从而减少副作用

- 选取适当长度的文章，创建用户词典之前与之后的分别翻译，检查翻译结果的区别。
- 频繁出现的一般名词，多数情况下会造成翻译质量的下降。检查到此种情况，请删除一般名词。

### 3. 创建专有名词时，检查翻译结果是否需要根据上下文而变化

- 用户词典无法通过上下文来改变翻译结果。因此当翻译结果需要根据上下文而变化时，最好避免创建该专用名词。
- 当然，可以通过创建多个用户词典，使用不同的配置文件来产生不同的翻译结果。

### 4. 无须创建也能获得较准确的翻译结果的单词，不建议创建用户词典

- 用户词典，不仅仅是匹配单词，匹配单词附近的译文也会受到影响。多数情况下，不创建用户词典的整体翻译质量较好，因此无须创建也能获得较准确的翻译结果的单词，请尽量避免创建用户词典（当然，为了防止根据文章不同而造成该单词翻译质量参差不齐而创建用户词典的情况也有，但正确翻译为多数情况的话最好避免创建该单词）。

## 4-3. 用户词典 使用窍门 (高级篇)



	原文	翻译结果
添加前	ペーパードライバー講習など、運転技能の向上を目指した研修を行います。	进行纸制驾驶员讲习等以提高驾驶技能为目标的研修。

「ペーパードライバー」应翻译为挂牌司机而非纸制驾驶员。通过创建词典来解决。

	原文	翻译结果
添加后	ペーパードライバー講習など、運転技能の向上を目指した研修を行います。	进行挂牌司机讲习等以提高驾驶技能为目标的研修。

Good

### 👉 推荐的创建注册方式

- 创建词典 “ペーパードライバー → 挂牌司机”  
[结果] 将 ペーパードライバー 解释成 挂牌司机 以外的可能性很低。(当然, 整体的翻译精度有可能因使用了用户词典而变差, 所以需要谨慎使用 [用户词典的特性](#))

### 👉 不推荐的创建注册方式

- 创建词典 “ペーパードライバ → 挂牌司机”  
[结果] 除非与原文里词语完全匹配, 否则不会反映 [用户词典使用窍门\(初级篇\) 3.理解会有无法正常反映的情况](#)  
[原因] ペーパードライバ 和 ペーパードライバー 是不同的词, 所以无法反映出来
- 创建词典 “ペーパー → 挂牌”  
[结果] 检查翻译结果是否需要根据上下文而变化 [用户词典使用窍门\(高级篇\) 3](#)  
[原因] 在许多情况下, ペーパー 被翻译为“纸”或者“纸制”, 因此强行添加进词典的话可能会对翻译结果产生意想不到的恶劣影响

## 4-3. 用户词典 使用窍门 (高级篇)

	原文	翻译结果
添加前	在宅勤務をきっかけにワンオペ育児をスタートさせることになりました。	以在家工作为契机，开始了一次手术育儿。

「ワンオペ育児」应翻译为单方育儿而非一次性手术育儿。通过创建词典来解决。

	原文	翻译结果
添加后	在宅勤務をきっかけにワンオペ育児をスタートさせることになりました。	以在家工作为契机，开始了单方育儿。

Good

### 👉 推荐的创建注册方式

- 创建词典“ワンオペ育児 → 单方育儿”

### 👉 不推荐的创建注册方式

- 创建词典“ワンオペ → 单方”

※请看[用户词典 使用窍门 \(初级篇\) 1.](#) 理解匹配的优先顺序，**优先匹配字数更多的会提高翻译质量。**

## 5-1. 什么是翻译记忆库?



翻译记忆库：是指通过成对地添加原文和译文，在输入的原文与添加的原文完全相同的情况，将与其对应的译文作为翻译结果输出的功能。

- 翻译记忆库适用于以下情形
  - 常用的句子
  - 章节标题或表格标题等
- 在翻译记忆库中，只要从句子的开头到结尾任一部分不一致，指定的译文就不会反映出来。

示例：私たちはガーラ湯沢に行きます。⇔ 我们要去GALA汤泽。

- 已添加到翻译记忆库的情况下，  
「もしみんなのスケジュールが空いていれば、私たちはガーラ湯沢に行きます。」  
这样的原文，后半部分就不会反映出指定的译文。
- 「私たちは  
ガーラ湯沢に行きます。」  
这样的中间插入换行的话同样不会反映出指定的译文。

## 5-2. 翻译库注册提示



### 1. 登录经常使用的句子

- 邮件的问候语、定型文
- 文档标题页脚上的语句等

原文	译文
一直承蒙您的关照。	Thank you for your support always.

### 2. 注册章节标题和表中的项目名称

- 论文或手册的章节标题

原文	译文
1.简介	1. Introduction

※在使用Microsoft Office的项目符号功能翻译文件时，只注册项目符号的文本部分。(在上述情况下，不包括1.)

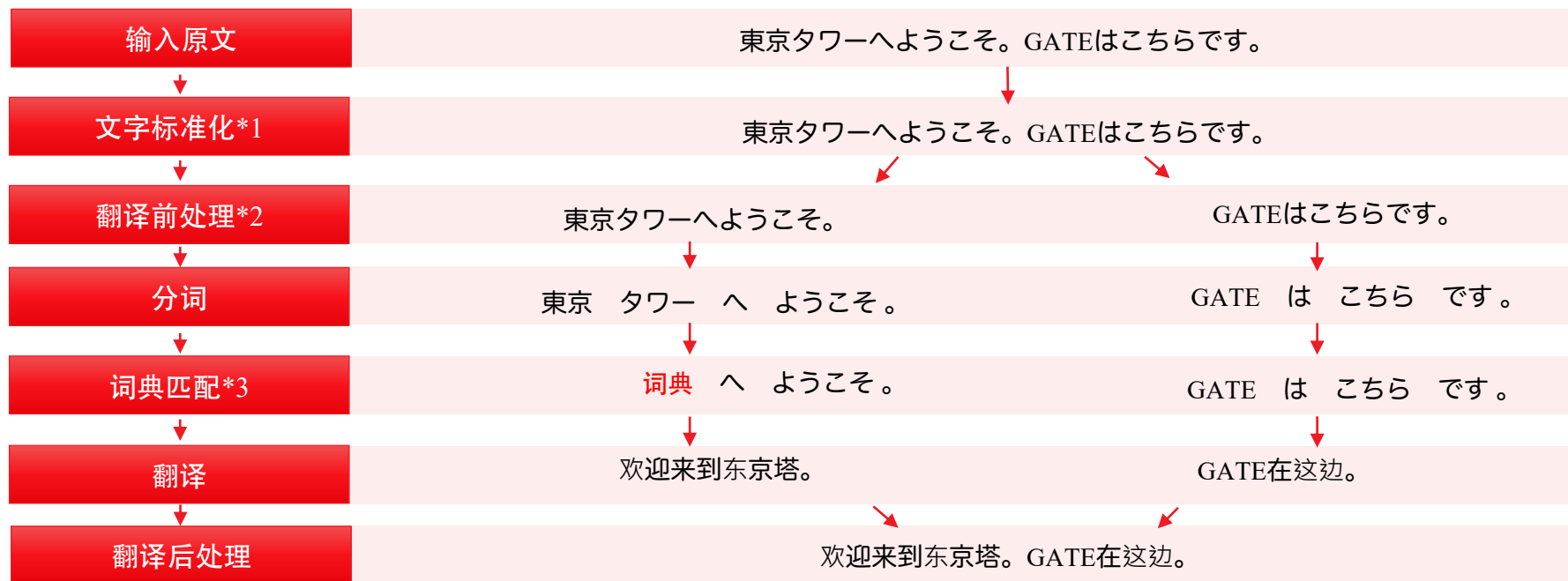
- 表中的项名称

品名	数量	单价	金额
XXX	1	100	100

在上面的表中，将项目名“品名”“数量”“单价”“金额”分别登录到翻译记忆库。

(只有在句子整体完全一致时，才能利用所登录的译文所反映的翻译记忆库的特性，将这些项目名称作为“句子”捕捉，并登录到翻译记忆库中。当然，它是“完全匹配”的，因此它反映了注册的内容。)

## 6. 参考资料 翻译的流程



- “分词”是指把句子分割成单词的过程。字典的匹配是在分词之后进行的。
- 在创建用户字典的建议3. 理解会有无法正常反映的情况的示例中, 「ガーラ湯沢駅」根据分词被分割为「ガーラ湯沢」和「駅」, 而「ガーラ」和「ガーラ湯沢」被看作是完全不同的词组, 因此即使在字典中添加了「ガーラ」, 也不会被匹配。然而, 如上述示例, 跨多个单词的词组是会被匹配的。(换言之, 添加的词组的开头和结尾必须是分词之后的词组的边界。)

\*1 文字标准化: 是指采用NFKC标准化, 将全角数字和字母转换为半角。

\*2 翻译预处理: 是指对句子进行分割等处理。

\*3 在用户字典中添加了 東京タワー ⇔ 东京塔 的情况