

みんなの自動翻訳@TexTra®

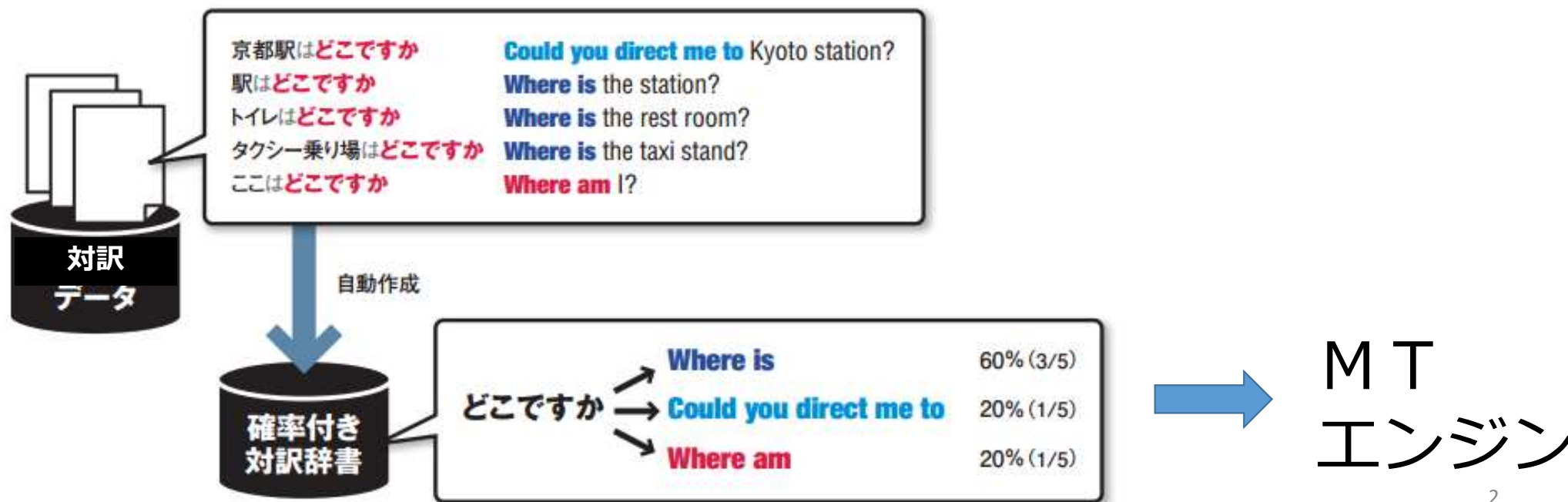


2016年6月

<https://mt-auto-minhon-mlt.ucri.jgn-x.jp/>

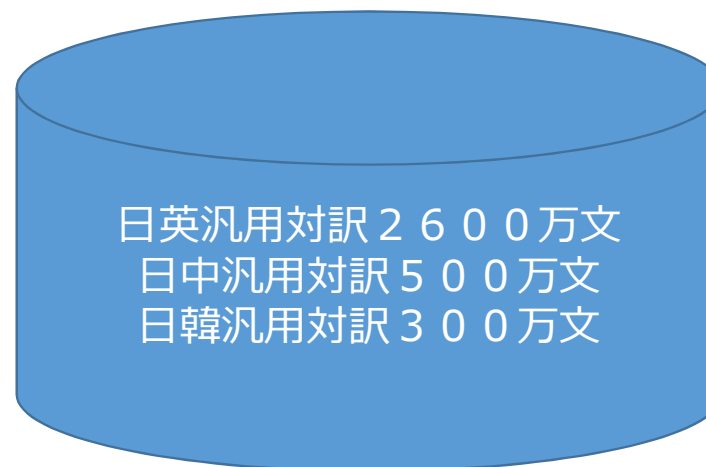
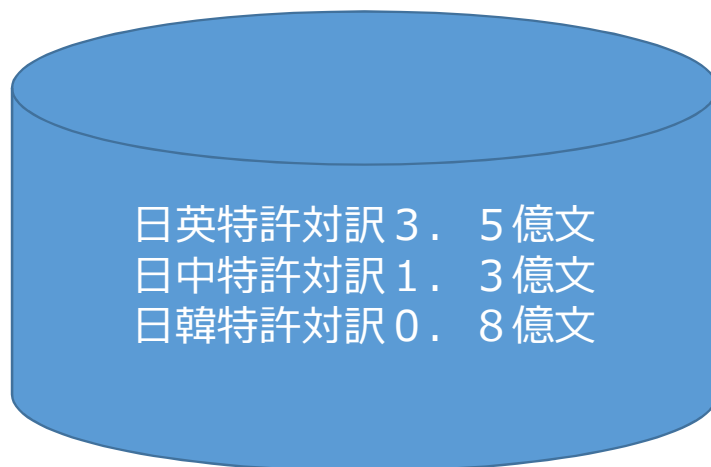
みんなの自動翻訳ができるまで①

- NICTの自動翻訳は統計的自動翻訳（SMT）です
- SMTは対訳データから自動的にMTエンジンを作ります



みんなの自動翻訳ができるまで②

- MTエンジンの作成には多量の対訳データが必要です
- NICTと特許庁は世界最大の特許対訳データを作成しました
- NICTでは今、汎用翻訳のための対訳データを集めています





みんなの自動翻訳ができるまで③

- MTエンジンにはアルゴリズムとソフトウェアも重要です
- NICTは、自動翻訳について多くの研究を公表し続けているとともに、研究成果を社会に還元しています

<http://www2.nict.go.jp/astrec-att/index.html>

- みんなの自動翻訳は社会で役立っています

TOPPAN

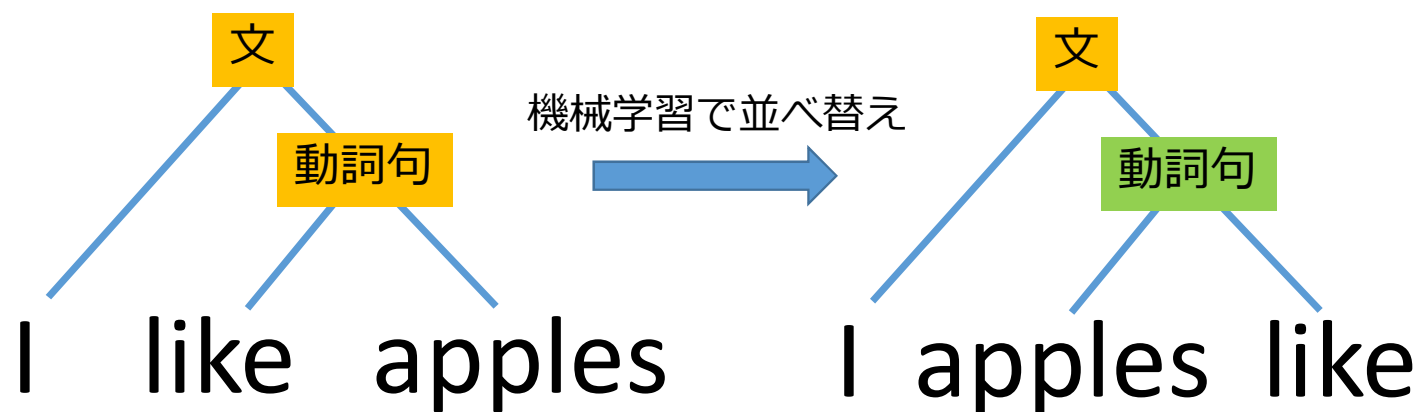


MiraiTranslate



長文翻訳エンジンの仕組み①

- 英日の長文翻訳では「語順の違い」と「訳語選択」が課題です
- 初めに、英語を日本語に似た語順に並べ替えます
- 英語の構文構造を調べて、機械学習で並べ替えます



特許2012-122866
特願2015-018281
特願2016-085262

長文翻訳エンジンの仕組み②

- 並べ替えられた英文は、フレーズごとに訳します
- フレーズの訳は、「確率付き対訳辞書」にあります
- 並べ替えの後なので、訳語選択はスムーズです

I → 私は

特許2008-145533

apples like → リンゴが好きです

長文翻訳エンジンの仕組み③

- 技術移転のために、研究成果を実用化しました
- みんなの自動翻訳は、コモディティサーバーで動きます
- 信頼できる環境下で、セキュアな翻訳環境を構築可能です



実用化

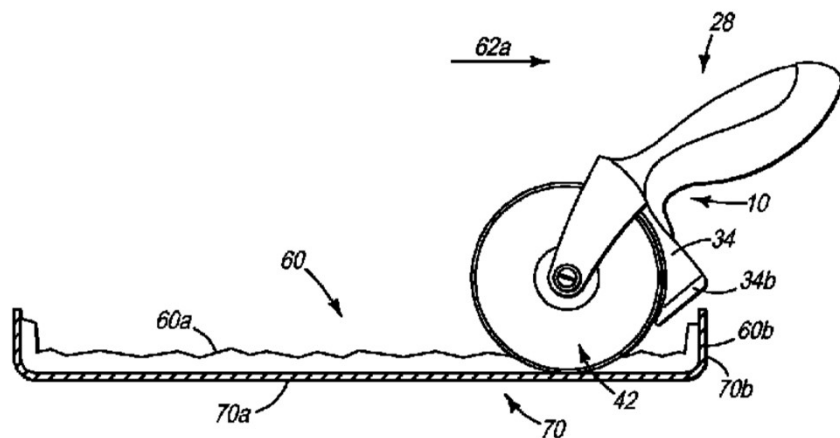


セキュアな翻訳環境

日英中韓汎用翻訳
日英中特許翻訳
日英中特許請求項翻訳

特許請求項翻訳エンジンの仕組み①

- 特許請求項は特許で一番重要です
- 特許請求項は特別なパターンを持つ超長文です
- 従来のSMTでは特許請求項の精度良い翻訳は不可能でした



A rotary cutter comprising:
 a handle;
 an arm extending from an end of the handle;
 a circular blade coupled to the arm;
 and a crust cutter member extending from the end of the handle,
 wherein an edge of the crust cutter member is sharpened and the crust cutter member extends in a plane that is substantially parallel to a centerline of the handle.

特許請求項翻訳エンジンの仕組み②

- N I C T はパターン翻訳を S M T に導入しました
- 最初にパターン解析をして、請求項の各部品を並び替えます
- 各部品は特許翻訳エンジンで翻訳します

部品名		文字列
前提部		An apparatus,
移行部		comprising:
本体部	構成要素	a pencil;
	構成要素	an eraser attached to the pencil; and
	構成要素	a light attached to the pencil.

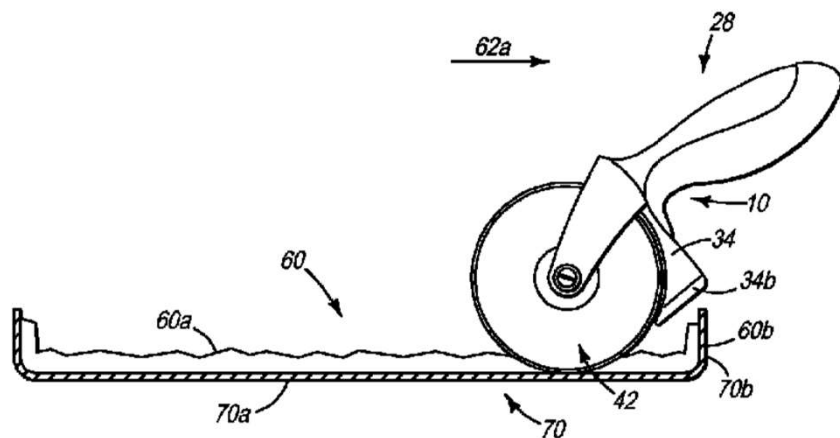
部品名		文字列
前提部		一种装置
移行部		包括：
本体部	構成要素	铅笔和；
	構成要素	安装在所述铅笔的橡皮擦以及；
	構成要素	安装在所述铅笔的灯泡。



部品名		文字列
本体部	構成要素	鉛筆と；
	構成要素	鉛筆に取り付けられた消しゴムと；
	構成要素	鉛筆に取り付けられたライトと
移行部		を備える
前提部		装置。

特許請求項翻訳エンジンの仕組み③

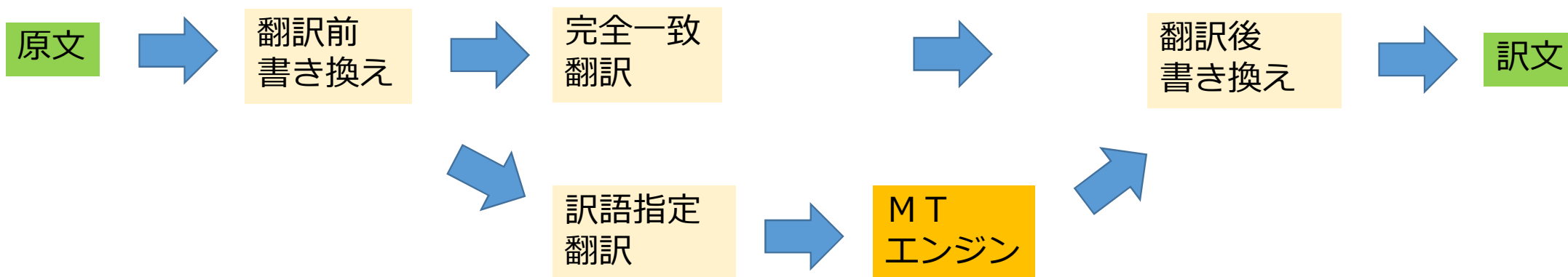
- 日英中の特許本文と特許請求項の翻訳に対応しています
- 英日特許翻訳は2億文、中日は1億文の対訳を活用しています
- 特許庁との協力により作成されたMTエンジンです



ハンドルと；
 前記ハンドルの前記端部から延びるアームと；
 前記アームに連結された円形のブレードと；
 前記ハンドルの前記端部から延在する外皮切断部材と；
 を備え、
 外殻部材の縁部は鋭利であり、皮の切断部材が、ハンドルの中心線に実質的に平行な平面に延在している、ロータリカッタ。

カスタマイズの仕組み①

- 新しい単語や新しい訳語が常に出現します
- 分野によっても翻訳は異なります
- NICTのMTエンジンはカスタマイズをサポートします



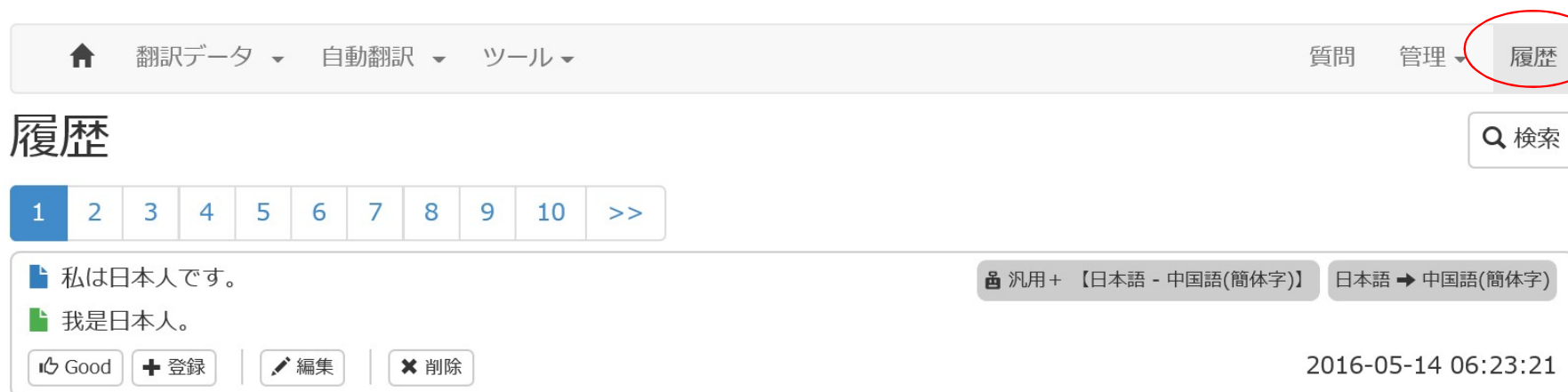
カスタマイズの仕組み②

- 「翻訳前書き換え」や「翻訳後書き換え」では、「はあ〜い」のような言い回しを「はい」に統一したりできます
- 「訳語指定」では、たとえば「情報通信研究機構」の訳語を「NICT」に特定できます

自動翻訳エンジン	翻訳
原文	私は 情報通信研究機構 に勤務して います 。
汎用+ 【日本語 - 英語】	I'm working at the National Institute of Information and Communications Technology .
MYMT	I'm working at the NICT.

カスタマイズの仕組み③

- 「完全一致翻訳」は登録している対訳と同じ入力にたいして、登録している訳文を出力します
- マニュアル翻訳等で特に有用です
- MTの履歴を修正してから対訳登録してください



The screenshot shows the NICT translation interface. At the top, there is a navigation bar with a home icon, '翻訳データ', '自動翻訳', and 'ツール' dropdown menus. On the right side of the navigation bar, there are links for '質問', '管理', and '履歴', with '履歴' circled in red. Below the navigation bar, the title '履歴' is displayed on the left, and a search box with '検索' is on the right. A pagination bar shows numbers 1 through 10, with '1' highlighted. Below the pagination bar, there is a translation entry. The source text is '私は日本人です。' and the target text is '我是日本人。'. The interface includes buttons for 'Good', '登録', '編集', and '削除'. The date and time '2016-05-14 06:23:21' are shown at the bottom right of the entry. A language selector at the top right of the entry shows '汎用+ 【日本語 - 中国語(簡体字)』 日本語 → 中国語(簡体字)'.

カスタマイズの仕組み④

- 対訳文を数万文以上保有している場合には、オリジナルMTの作成をお試しく下さい
- 自分のMTは自分の文章を精度よく翻訳できます

🏠 翻訳データ ▾ 自動翻訳 ▾ ツール ▾ 質問 管理 ▾ 履歴

オリジナル自動翻訳一覧 / オリジナル自動翻訳登録

オリジナル自動翻訳登録 ヘルプ

名前

学習タイプ

言語方向

- 英語 → 日本語
- 英語 → 韓国語
- 英語 → 中国語(簡体字)
- 英語 → 中国語(繁体字)
- 日本語 → 英語
- 日本語 → 韓国語

ホームページ翻訳①

- URLからの機械翻訳が翻訳エディタで表示されます
- 翻訳エディタで機械翻訳を修正します



The screenshot shows a web browser window with the following content:

- Browser address bar: 戻る, 保存, ファイル, 一括, 元サイト, 表示設定, ヘルプ, WYSIWYG
- Logo: 経済産業省 (Ministry of Economy, Trade and Industry)
- Breadcrumb: > Announcements > News Releases > Fiscal 2014 lists > Patent Office and the National Institute of Information and Communications Technology (NICT) agreed to cooperation in machine translation of patent documents
- Main text: **Patent Office and the National Institute of Information and Communications Technology (NICT) agreed to cooperation in machine translation of patent documents**
- Section header: **Outline of the case**
- Text: Patent Office and the National Institute of Information and
- Footer: なし, 文章数 0 / 58, 文字数 2403 | 7087, 185%

ホームページ翻訳②

- 翻訳を修正するときには、辞書引きなどの支援があります
- Web検索もできるので便利です



産業界
of Economy, Trade and Industry

ments ▶ News
s Technology

Office and
ation and Communications
agreed to cooperate in machine translation
it documents

特許庁と独立行政法人情報通信研究機構（NICT）は特許文献の機械翻訳に関する協力を合意しました

Patent Office and the Agency of Information and Communications Technology (NICT) agreed to cooperate in machine translation of patent documents

行政

行政

administration

execution

executive

government

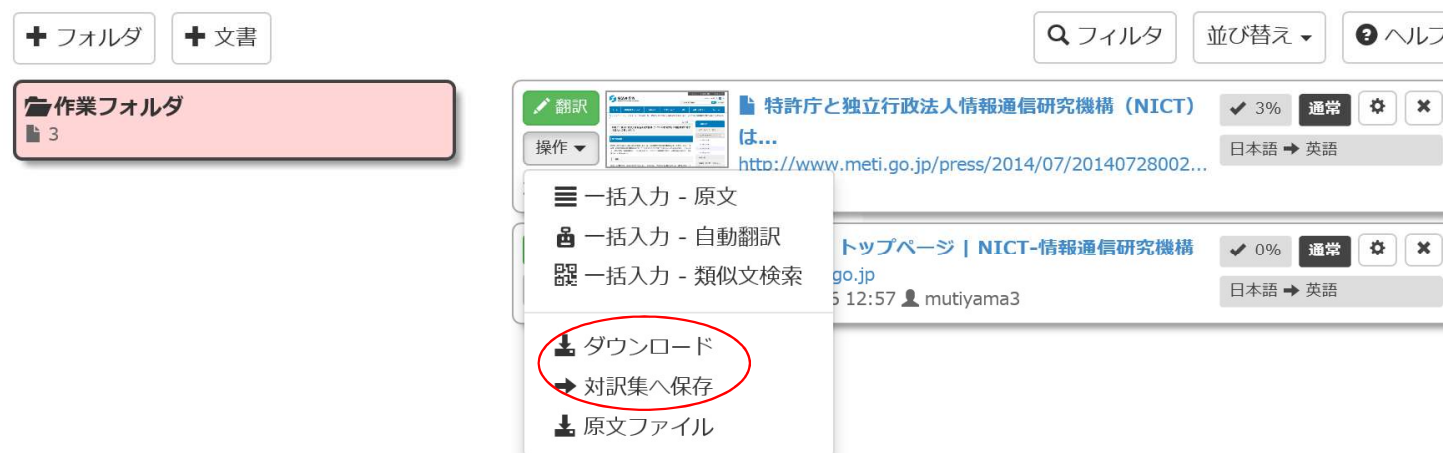
public administration

更新 コミット 2

TM 履歴

ホームページ翻訳③

- 翻訳結果はHTMLとしてダウンロードできます
- 翻訳結果を後で活用するには、対訳集へ保存してください
- 翻訳メモリとして活用可能です。MTの性能も向上します



みんなで翻訳①

- 共同翻訳者を登録してみんなで翻訳ができます
- ツールバーの「管理」から「ユーザ」を選びます
- 登録されたユーザは同じ文書を編集できます

🏠 翻訳データ ▾ 自動翻訳 ▾ ツール ▾ 質問 管理 ▾ 履歴

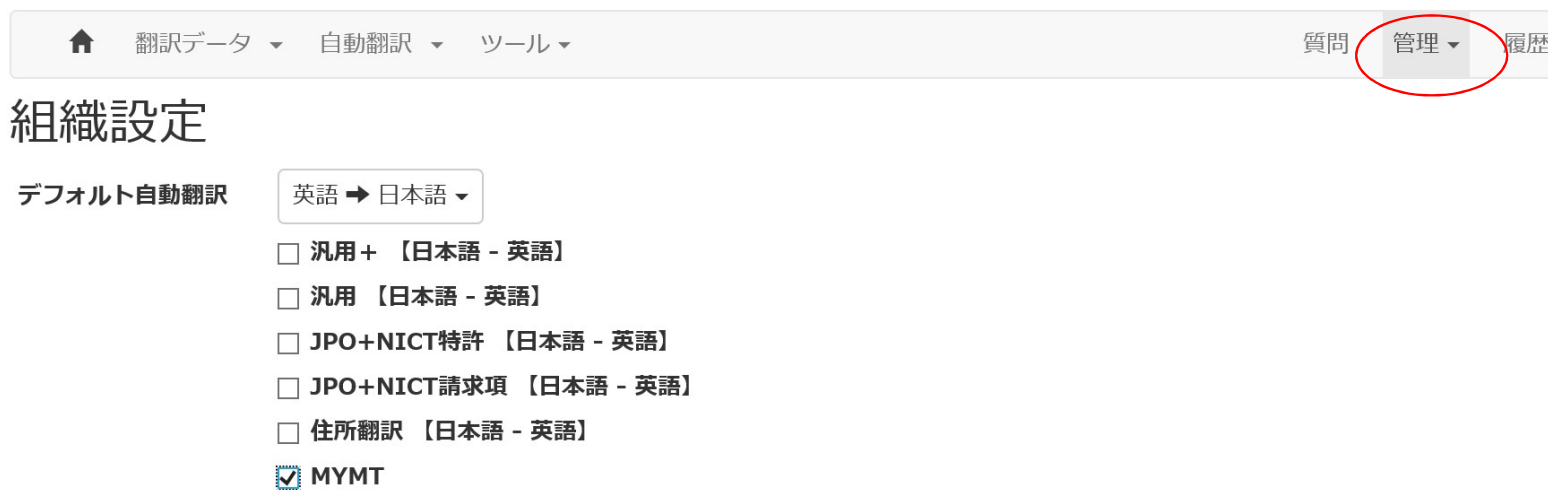
ユーザー一覧

+ ユーザー登録

組織	名前	タイプ	ログ
★ [redacted]	 [redacted]	組織管理者	ログ
★ [redacted]	 [redacted]	一般ユーザー	ログ

みんなで翻訳②

- カスタムMTをみんなで共有できます
- ツールバーの「管理」から「組織設定」を選びます
- カスタムMTをデフォルト自動翻訳に設定してください



Home 翻訳データ ▼ 自動翻訳 ▼ ツール ▼ 質問 管理 ▼ 履歴

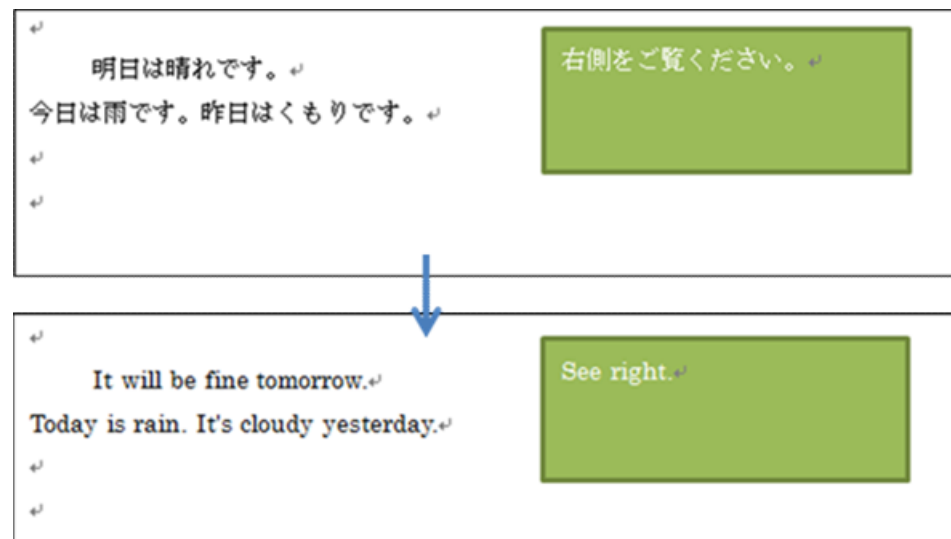
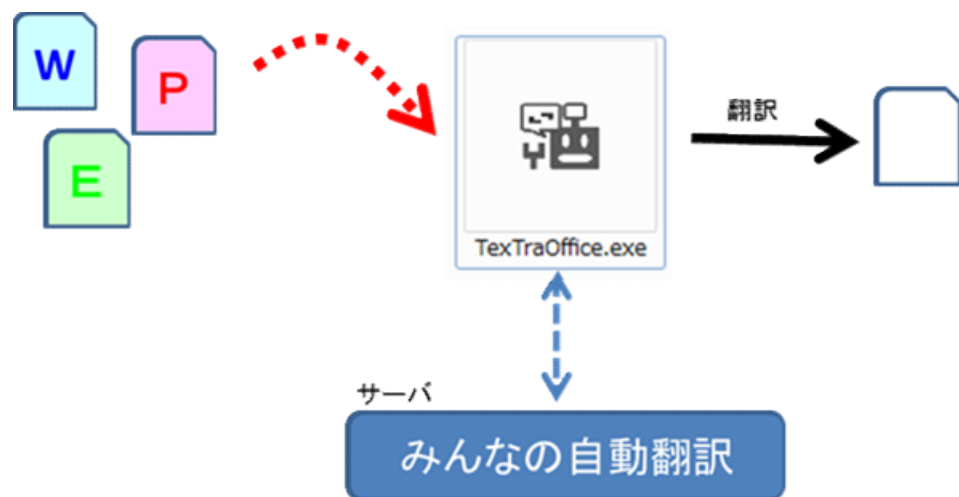
組織設定

デフォルト自動翻訳 英語 → 日本語 ▼

- 汎用+ 【日本語 - 英語】
- 汎用 【日本語 - 英語】
- JPO+NICT特許 【日本語 - 英語】
- JPO+NICT請求項 【日本語 - 英語】
- 住所翻訳 【日本語 - 英語】
- MYMT

その他のツール①

- 「TexTraクライアントツール TexTraOffice」はWord、Excel、PowerPointファイルの中身を翻訳文に置き換えます



その他のツール②

- 「単言語用語抽出」は、「職業翻訳」に必須の機能です
- 翻訳対象の文書を登録して、しばらくお待ちください
- このパワポからは、下の用語と訳語（自動翻訳）が出てきます

16	翻訳	Translation
11	自動翻訳	Automatic translation
7	カスタマイズ	Customize
5	MTエンジン	MT engine
4	長文翻訳エンジン	Long translation engine
4	特許請求項翻訳エンジン	Patent claims translation engine
4	特許請求項	Patent claim
4	特許	Patent
4	書き換え	Rewriting
4	レベル	Level.

その他のツール③

- 「TexTraクライアントツール Word Addin」はMicrosoft Word上での翻訳作業をサポートするWordアドインです
- 翻訳者さんが長年のWord上での翻訳作業で獲得したメリットを損なうこと無く、翻訳データベースとネット連携して、便利な機能を提供します。用語抽出や翻訳メモリも連携できます。



今後の予定

- 多言語に拡張します

MT種別	原言語	訳言語
汎用	日英	日英中韓、アラビア、ドイツ、スペイン、フランス、インドネシア、ブラジルポルトガル、タイ、ベトナム、ミャンマー
特許	日英	日英中韓独仏

- ScrumSourcingを初めとするクラウド翻訳を開始します

ScrumSourcingは立教大学トニー・ハートレー教授、豊橋技術科学大学井佐原均教授、NICTが中心となって、ラグビーワールドカップ2019の翻訳を支援するプロジェクトです