＜論文＞

法令用翻訳メモリデータベースシステムの開発

The Development of Translation Memory Database System
for Law Translation

関根　康弘
（立教大学大学院）

Abstract

This paper introduces the development of “Japanese Law Translation Memory”: a translation memory database system for translating Japanese laws and regulations. This system provides the translation memory via the Internet. Users such as translators can download the entire content of the database as TMX and CSV format for use with computer assisted translation tools. This system also has a search function that enables users to find similar translated source sentences and their corresponding translations in the database. This paper aims to show the background and purpose of the development, overview the database and functions of the system and discuss about problems and tasks for further development.

1.　はじめに

近年、法令翻訳のニーズが高まっている。このニーズにこたえる形で 2009 年 4 月に「法令外国語訳データベースシステム（英語名：Japanese Law Translation、以下 JLT）」が法務省によって公開され、法令翻訳とそれに関連するさまざまな情報が提供されるようになった。筆者は名古屋大学に協力し、このシステムの設計と開発の携わり、現在もシステム及びデータの管理運用に携わっている。

JLT が公開されるまでは、法令の翻訳は各所管の府省庁によって個別に行われており、文書の書式やデータの形式も統一されていなかった。このため、利用者にとってはどこに必要とするデータがあるのかわかりづらく、利用しにくいものであった。JLT が公開されてからは翻訳データが一か所にまとめられ、文書の書式とデータの形式も統一され、利便性が高まった。しかし、JLT の公開後も翻訳の作業自体は、従来どおり各所管府省庁によって個別に行われており、このことが統一性のない翻訳や品質のばらつきの原因となっている。

この問題を解決するため、筆者は法令用翻訳メモリデータベースを開発し、2011 年 9 月より試験的に公開を開始した。これまでに各所管府省庁によって翻訳されたすべてのデータを翻訳メモリとしてひとつにまとめ、これを一般公開することにより、分散した法令翻訳のリソースを所管府省庁横断的に共有できるようになった。翻訳メモリデータベースは現在 JLT で公開されている法令デ

ータを元に構築した。このシステムでは 2012 年 3 月 13 日現在、259 本の法令データを元にして作成したユニット数 147,119 の翻訳メモリを提供している。

本稿ではこの法令用翻訳メモリデータベースシステムをテーマに、この研究の背景について（次項）、システムの概要や機能について（第 3 項）を述べたのち、現システムの課題点を整理して、システムの改善について（第 4 項）も検討する。

本稿は 2011 年 9 月に行われた第 12 回日本通訳翻訳学会年次大会での口頭発表の内容をまとめ、それに加筆したものである。

2. 研究の背景

2.1. 法令翻訳のニーズ

近年、経済のグローバル化、法整備支援活動、在留外国人への法情報の提供などの必要性から法令翻訳のニーズが高まっている。グローバル化された経済の中で、日本に対する外国からの投資意欲をいっそう高めるためには、ビジネスの基礎となる法についての情報を幅広く提供することが必要だという認識は、経済界や官庁に幅広くみられる（外山, 2004）。外国からの対日投資を促進するには日本法を外国語訳（とりわけ英訳）で提供することが不可欠である。また、法整備支援活動においても法令の翻訳が必要とされる。開発途上国や体制移行国における経済発展のための市場経済体制の確立には法整備支援が必要である。法整備支援は昨今、先進諸国によってさかんに行われており、日本政府も国策として 1990 年代からアジア諸国を中心として法整備支援活動を行っている。さらに、在留外国人への法情報の提供という点でも法令翻訳が必要とされる。日本の外国人人口は 2009 年まで 20 年にわたって増加し続けている。欧米諸国に比べると人口比は 1.8%でそれほど高くないが、外国人人口の約 3 分の 2 が定住者と言われ、この数字から見ると世界的な移民受け入れ国であるということができる。それにもかかわらず、外国人に対する社会システムや行政サービスが充実しているとは言い難い。在留外国人にとって公正な社会を実現するには法令翻訳は不可欠である。

2.2. 法令外国語訳データベースシステム

このようなさまざまな理由から法令翻訳が必要とされている。この高まるニーズにこたえる形で 2009 年 4 月法務省によって「法令外国語訳データベースシステム（JLT）」が公開された。公開から今日まで 1 日平均で 11 万件以上のアクセス、毎月 2 万以上のアクセス元のドメイン数を記録しており、これらの数字が法令翻訳のニーズの高さをよく表している。この JLT は名古屋大学によって設計・開発が行われた。筆者は名古屋大学に協力する形で設計・開発に携わり、現在もシステムメンテナンスやデータの追加・更新などの作業を担当している。

図 1　法令外国語訳データベースシステム(トップページ)

JLT が公開されてから約 2 年半以上が経ったが、これまでこの高いニーズを保ちつつ大きな問題もなく安定的な稼働を続けている。JLT が公開されるまでは法令の翻訳は各所管の府省庁がそれぞれ別々に提供していたため、必要なデータの入手が非常に困難であった。また、提供する文書の書式やファイル形式もばらばらで利用しにくいものであった。しかし、JLT の公開によって翻訳データは一か所にまとめられて横断的な検索が可能となり、データも PDF、MS-WORD、テキスト、XML などの統一された各種形式で提供されるようになり、利便性が大きく向上した。

しかし、JLT が公開されてからも翻訳作業自体は一元化されておらず、従来どおり各所管の府省庁によって個別に行われている。このことにより、翻訳のリソースやノウハウが分散している状態になっており、統一性のない翻訳や品質のばらつきの要因となっている。また、各所管の府省庁は翻訳作業を外部委託しているケースが多く、通常は入札によって業者が決定される。しかし、法令翻訳のスキルを測るような入札資格の基準がないため、金額の安さのみで決定される場合が多いようである。これも品質の低下を招く原因となっていると推測される。

このように JLT の開発・運用に携わるうち、法令翻訳の問題点がいくつか見えてきた。それらを次節にまとめる。

2.3. 法令翻訳の問題点

法令翻訳の持つ問題は大きく分けて量の問題と質の問題に分けることができる。量の問題とは、一言でいうと、必要な法令の翻訳データがあるか、ということである。現在 JLT では 259 本(改正バージョン違いを別々にカウントすると 284 本)の法令の翻訳データが公開されており、(2012 年 3 月 13 日現在)この数は日々増え続けている。しかし、法令は全部で約 7000～8000 本あると言われているので、ユーザからの高いニーズを満たすには十分であるとは言い難い。また、府省庁ご

とに法令の翻訳計画 [1] が公開されているが、計画の遅れが目立つようになっている。このような理由から一般ユーザからの翻訳のリクエストが頻繁に寄せられている。

質の問題とは、訳文が原文の意味を正確に表し、理解しやすいかということである。JLT では、各翻訳担当者から提出されるデータには訳抜け、ダブルスペース、スペルミスなどのケアレスミスが多く、一般ユーザから誤訳の指摘を受けることもたびたびある。また、同一の原文に異なる訳文が当てられているケースが多くみられ、利用者の理解の妨げになっている。以下は不統一訳の一例である。

この法律は、公布の日から施行する。
This Act shall be enforced from the date of promulgation.
This Act shall be enforced from the day of promulgation
This Act shall come into effect as from the date of promulgation.
This Act shall come into effect as from the date of promulgation;
This Act shall come into effect as from the day of promulgation.
This Act shall come into effect as from the day of promulgation;
This Act shall come into effect as of the date of promulgation;
This Act shall come into effect as of the day of its promulgation.
This Act shall come into effect as of the day of its promulgation;
This Act shall come into effect as of the day of promulgation.
This Act shall come into effect as of the day of promulgation;
This Act shall come into effect on the day of promulgation.
This Act shall come into effect on the day of promulgation;
This Act shall come into force as from the date of its promulgation.
This Act shall come into force as from the date of promulgation,
This Act shall come into force as from the date of promulgation.
This Act shall come into force as from the date of promulgation;
This Act shall come into force as from the day of its promulgation.
This Act shall come into force as from the day of promulgation,
This Act shall come into force as from the day of promulgation.
This Act shall come into force as from the day of promulgation:
This Act shall come into force as from the day of promulgation;
This Act shall come into force as of the date of its promulgation.
This Act shall come into force as of the date of its promulgation;
This Act shall come into force as of the date of promulgation.
This Act shall come into force as of the day of its promulgation.
This Act shall come into force as of the day of promulgation.
This Act shall come into force from the day of promulgation.
This Act shall enter into force on the day of its promulgation,
This Act shall enter into force on the day of promulgation.

図 2　不統一訳の例

「この法律は、公布の日から施行する。」という表現は法律において非常によく用いられる表現であり、このような定型的な文言は統一して訳されるべきであるが、JLT ではこの原文に対して 30 とおり以上の訳文が存在する。さらに極端な例になると、同一の文に対する訳として 120 とおり以上の訳文が存在した。このような訳のばらつきは将来的には法令英訳専門家によって検討され、統一されることを期待したい。そして、その統一された訳を取り入れた翻訳メモリを本システムで提供することができれば、新たに行われる翻訳でも訳を統一することが可能となる。

2.4. 翻訳メモリの公開

前節で挙げた問題を解決するため、翻訳メモリに着目した。翻訳メモリを公開することによって次のようなことを期待している。

まずは、作業の効率化が挙げられる。作業の効率化はコストの削減と作業期間の短縮につながる。コストが削減できれば、余った予算をほかの法令の翻訳に充てることができるため、公開法令の増加につながる。また、作業期間が短縮できれば、スケジュールどおりの公開が可能となる。

次に品質の向上が挙げられる。翻訳メモリの使用により、必然的に訳が統一される。また、公開されているデータはさまざまなチェック・修正がなされているので、これらを用いることで誤訳やケアレスミスの防止につながる。

さらに、メモリの公開によって、法令の翻訳がより多くの人、とりわけ、法律家、翻訳者、研究者など専門家の目に触れることにつながり、これらの利用者からのフィードバックを得ることができれば翻訳メモリ自体の品質を高めるサイクルができると期待される。また、JLT で新たな法令の翻訳が公開され続けている限り、翻訳メモリのデータ量も増え続けることになるため、下図のような追加／修正のサイクルによって持続的な質の向上と量の拡充が可能となる。

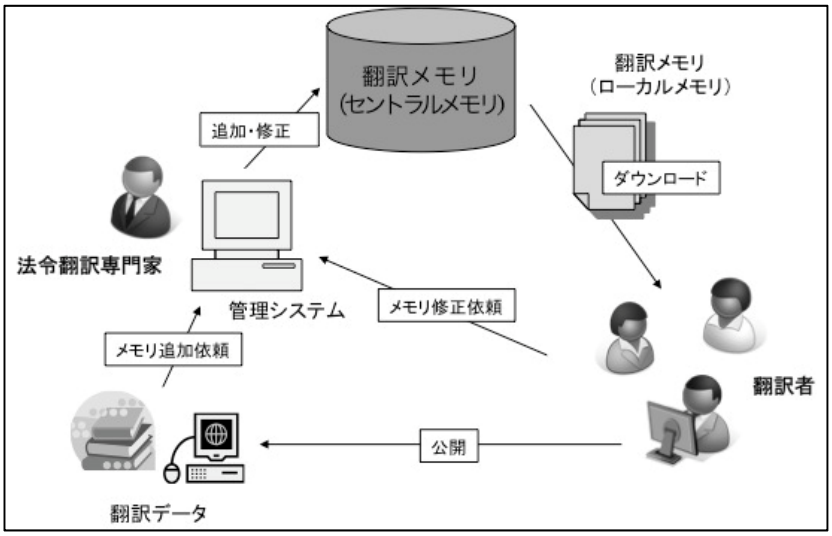

図 3 翻訳メモリの追加／修正のサイクル

翻訳メモリのマイナス面として、誤訳の伝播ということがよく言われるが、誤訳が見つかったら直せばよいのである。翻訳メモリのように、文単位ですべてのデータを一元的に管理しておけば、誤訳であると気が付いたときに全データの中から該当する部分を検索し一括して直したり、出典のデータを特定し、まとめて修正したりすることも技術的には可能である。

3. 法令用翻訳メモリデータベースシステム

翻訳メモリは TMX[2] 形式などでファイル（ローカルメモリ）をダウンロードする形で利用者に提供するが、おおもとのデータ（セントラルメモリ）は一元的に管理される必要がある。これは利用者などからのフィードバックによる修正や新たなデータの追加などによって随時、修正、拡充が発生し、これを一元的に行うためである。このセントラルメモリでは原文と訳文だけでなく、出典の法令名、法令番号、条項番号、登録日、登録者、修正履歴などさまざまなメタ情報を管理する。このような

大まかな仕様は EU の翻訳総局で用いられているシステムに習ったものである。

3.1. EU の取組み

EU では早くから翻訳の効率化のための IT の導入がなされており、本研究において学ぶべき点が多い。EU には 23 の公用語があり、すべての法令はソース言語から 22 の言語へ翻訳されるため大量の翻訳（170 万ページ/2009 年）をこなす必要があり、さらにどの言語バージョンも法的効力を持つため、高い品質も要求される。EU 翻訳総局ではさまざまな管理システム、翻訳支援ツールを駆使して、この高度な要求にこたえている。このツール類はワークフローの中でお互いに連携し、作業の効率化と品質の管理に役立っている。このシステム／ツール類の中でもっとも中心的な役割を果たしているのが EURAMIS[3] という翻訳メモリ管理システムである。本研究の目的を一言でいうと、この EURAMIS の日本版を作ることであるとも言える。

3.2. Japanese Law Translation Memory

筆者は EU における EURAMIS に相当する翻訳メモリデータベース（セントラルメモリ）として「Japanese Law Translation Memory[4]」を開発し、2011 年 9 月より試験的な公開を開始した。安定的な公開のため、筆者の所属する企業からインフラ面でのサポートを受けている。なお、システムの開発、運用はボランティア（非営利）で行われている。ユーザは誰でもインターネットを介してシステムへアクセスすることができ、すべて無料で利用することができる。

システムには検索機能とダウンロード機能があり、ユーザはデータベースから類似文を検索したり、ダウンロードしたデータを翻訳支援ツールで読み込んだりして利用することができる。

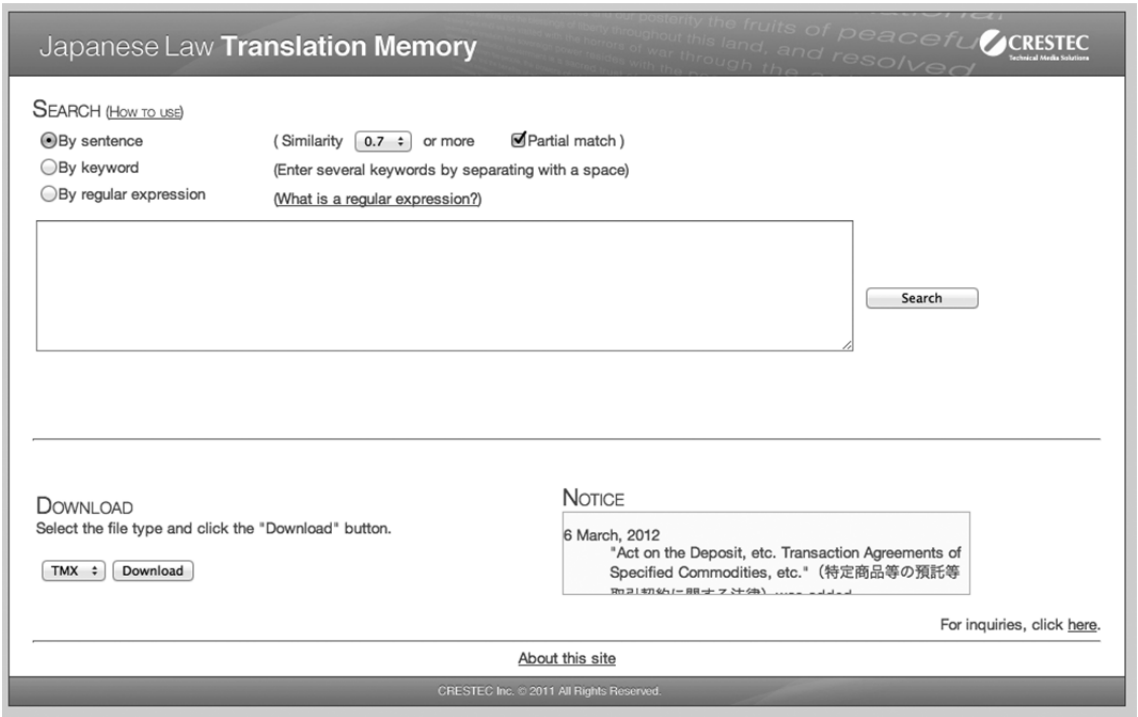

図 4　Japanese Law Translation Memory（トップページ）

3.2.1. データベースの構築

データベースの構築には現在 JLT で公開されている XML[5] を用いた。XML からテキストノード[6] の単位で文を抜き出し、日英ペアにして 1 レコードとなるようにした。日英のペアを作る際には次の手順のように XPath[7] を使用した。

1) 日本語 XML からテキストノードとその XPath を取り出す。
2) 取り出した XPath を用いて英語 XML から文書構造的に同じ場所に位置する英語のテキストを取り出す。
3) 日英のテキストをペアにしてデータベースに蓄積する。
4) 上記の 1)～3)をすべての XML のすべてのテキストノードに対して行う。

このような手順をとった理由は、原文と訳文のアラインメントをできるだけ正確に行うためである。元データの XML の精度は 100%ではなく、またそうであったとしても日本語で 2 文になるものが英語で 1 文というようなケースがあるため、テキストノードだけを抜き出して順番にマッチングさせただけでは、アラインメントにずれが生じる危険が高いためである。

2012 年 3 月 13 日現在、259 本の法令[8] によってレコード数 276,597、原文のことなり数 129,120、原文−訳文ペアのことなり数（翻訳メモリのユニット数）147,119 のデータベースが構成されている。

3.2.2 機能

ユーザは翻訳メモリを TMX または CSV 形式でダウンロードし、SDL Trados のような翻訳支援ツールで読み込んで使用することができる。さらに、このような翻訳支援ツールを持っていない法律家や研究者のために、ツールなしでも翻訳メモリを利用できるよう、検索機能も設けた。検索機能には「類似文検索」、「キーワード検索」、「正規表現検索」の 3 つがある。ユーザはこの 3 つの検索方法を用いてデータベースから検索をすることができる。

翻訳者が翻訳メモリを用いる場合、おもに翻訳対象文に一致する原文または類似する原文を翻訳メモリの中から検索し、その訳文を用いて翻訳を行う。本システムにおいても類似検索が翻訳支援機能のメインとなる。本システムでの類似文検索の検索結果は以下のように表示される。

図 5 類似文検索の結果表示

検索結果は類似度の高いものから順番に並べられ、類似度が同じ場合は出典の数が多い順番に並べられる。類似文の検索は自然言語処理などの分野からさまざまな方法が提案されているが、本システムでは竹中・若尾（2011）の方法を参考にした。竹中・若尾（2011）は計算機によって異なる 2 つの自治体の対応する条例の条文の対応付けの実験を行った。この対応付けは、2 つの文の間の類似度を計算し、もっとも類似するものを機械的に選び出す方法で行われた。96 の類似度計算方法で実験を行った結果、法律家の行った対応付けにもっとも近かったのは条文から漢字のみを抜き出し、文字アラインメントで類似度を計算する方法であった。筆者はこの計算方法が最適であると考え、本システムの類似文検索でも採用した。しかし、この計算方法は文の総当たりで計算するため、検索時間がかかりすぎるという欠点がある。そこで、あらかじめある程度の絞り込みを行った後に類似度を計算するようにした。類似文の絞り込みには岡崎・辻井 (2011)のアルゴリズムを採用し、処理を高速化した。

類似文検索のほかキーワード検索と、正規表現検索の機能も設けた。キーワード検索は指定したキーワードを含む文を検索するという一般的なものである。正規表現[9]検索では文のパターンを指定して検索することができる。

4. システムの課題点と改善の検討

法令翻訳におけるさまざまな問題を解決するため、筆者は Japanese Law Translation Memory を開発し、2011 年の 9 月より、インターネットでの公開を開始した。システム公開から随時データの追加を行い、プログラムの改修なども加えてきた結果、アクセス数も徐々に増えてきており、ある程度は翻訳者に役立つシステムとなっているのではないかと思われる。しかし、利用者からのフィードバックや既存翻訳支援システムとの比較、産業翻訳の実務との比較などから考えると、本システムにはまだ改善すべき点が多い。以下に現時点での 3 つの課題を挙げる。

第 1 の課題として、翻訳メモリのサイズが挙げられる。先行する研究[10]によって、法令翻訳においてある程度翻訳メモリが有効であるということがわかった。しかし、法令によってはメモリから類似文がほとんど見つからないというものもあるため、初めてシステムを利用したユーザがそのような法令をもとに本システムで検索を行った場合、何も結果が得られないことになってしまう。初めての使用時に複数回の検索を行い、一度も類似文が得られなかった場合、利用者に「このシステムは役に立たない」というイメージを与えかねず、そのようなイメージを与えてしまった場合、次回以降使ってもらえなくなる可能性がある。このような事態を避けるため、JLT で公開されている翻訳データ以外でデータベースを拡充する方法を検討したい。また、類似文が見つからなかった場合にも、検索文に用いられている用語に「法令用語日英標準対訳辞書[11]」の内容を表示したり、機械翻訳の結果を表示したりといった機能を追加することを検討している。

第 2 の課題としては、機能の充実が挙げられる。翻訳メモリ自体は産業翻訳では長く用いられてきており、SDL Trados のような洗練された翻訳支援ソフトウェアがいくつも存在する。このような洗練されたソフトウェアと比較すると、本システムの機能は非常に貧弱である。セントラルメモリとしてウェブで翻訳メモリを一元管理するという点では、本システムはそのほかの翻訳支援ツールとは目的が異なるが、すべての利用者が翻訳支援ツールを所有しているわけではないため、検索機能

を中心とした翻訳支援ツールとしても機能の充実をはかりたいと考えている。標準対訳辞書との連携やメモリ追加／修正のサイクルなど、法令に特化したシステムであるという点やデータが一元的に管理されているという点など、本システムの特徴を生かしたシステムの改善を考えて行きたい。

第 3 の課題は本研究開発の成果をどのように評価し、システムの改善につなげるかという点である。これは筆者自身が翻訳者でないため、システムの利便性や翻訳の品質についての評価ができないためである。本研究を行っていくにあたって、翻訳の品質の評価や実務家によるシステムの評価が不可欠であるが、筆者は法令翻訳の専門家ではなく、翻訳者でもないためこれができない。アンケートフォームを設置するなどして、システム利用者からのフィードバックを得ると同時に、法令翻訳の実務家や専門家の協力を得るなどして、法令翻訳の効率化と品質の向上に貢献できるシステムになるよう、各方面の方々から協力を得ながら本研究開発の評価を行い、システムの改善につなげたい。

.....

【謝辞】

本システムの開発は名古屋大学大学院付属法情報研究センターのこれまでの活動の成果が土台となっており、筆者が同センターの研究活動を通して着想を得たものである。同大学教授であり同センター長の松浦先生にはさまざまな研究会や国際会議への参加の機会を与えていただいた。また、同大学情報科学研究科の外山先生ならびに小川先生にはシステム開発にご助言をいただいた。ここに深謝の意を表する。本システムのインターネット公開にあたっては株式会社クレステックの情報インフラを利用し、web 開発には同社ソリューション事業部情報研究開発課の大谷氏にご助言をいただいた。ここに深謝の意を表する。本研究の口頭発表ならびに本稿執筆にあたっては立教大学異文化コミュニケーション研究科長沼先生および武田先生にご指導いただいた。ここに感謝の意を表する。

本研究の一部は人工知能研究振興財団の平成 22 年度研究助成金によった。

.....

【著者紹介】

立教大学大学院異文化コミュニケーション研究科修士課程、株式会社クレステック情報研究開発課スーパーバイザー、名古屋大学大学院附属法情報研究センター研究員

連絡先：11vt029r@rikkyo.ac.jp

.....

【注】

1) http://www.japaneselawtranslation.go.jp/rel_info/rel_info_trans?re=01
2) XML 形式の翻訳メモリの標準仕様
3) 集積された巨大な多言語翻訳メモリ（Central Memory）を管理するシステム。翻訳者はこのデータベースを直接利用するのではなく、翻訳対象の文書に最適化された翻訳メモリ（Local Memory）をダウンロードして翻訳に用いる。Central Memory の編集を行うインタフェースも用意されており、翻訳者からのフィードバックをもとに、定められた管理者（Sentence Manager）に

よって修正などが行われる。翻訳メモリの管理のほか、翻訳前の原文の下処理、関連資料の検索などの機能もある。

4) http://itrd.crestec.co.jp/transmemoryweb/

5) 文書のパーツに意味的なタグをつけ、構造化したもの

6) 1 つのタグに納められたひとまとまりのテキスト

7) XML 文書の特定の部分を指定する言語構文

8) 改正バージョン違いを別々にカウントすると 280 本

9) いくつかの文字列をひとつの形式で表現するための表現方法。
例）*この(法律|政令)において[、]?「.+」とは、*
*(法律|政令)*は法律または政令を表し、*[、]?*は読点が 0 または 1 個出現すること（=あってもなくてもよい）、*「.+」*は鍵括弧の中に任意の文字が 1 つ以上出現するということを表している。

10) 関根康弘・齋藤大地・小川泰弘・外山勝彦・松浦好治（2010）

11) 内閣に設置された「法令外国語訳推進のための基盤整備に関する関係省庁連絡会議」の下に置かれた「実施推進検討会議」において検討が行われ、その下に設けられた学者及び弁護士からなる「作業部会」によって作成された辞書。関係府省が法令を英訳する場合には、この辞書に準拠している。

【参考文献】

松浦好治（2009）「法令外国語訳プロジェクトの意義―日本法・法制度の国際通用性」『ジュリスト』2009 年 4 月 15 日号, 2-7 頁, 有斐閣.

岡崎直観・辻井潤一（2011）「集合間類似度に対する簡潔かつ高速な類似文字列検索アルゴリズム」『自然言語処理』Vol. 18, No. 2, 89-118 頁, 2011 年 6 月号

関根康弘・齋藤大地・小川泰弘・外山勝彦・松浦好治（2010）「法令翻訳における翻訳メモリの有効性」『平成 22 年度電気関係学会東海支部連合大会講演論文集』F1-1, 1 頁

竹中要一・若尾岳志（2011）「地方自治体の例規比較に用いる条文対応表の自動生成」言語処理学会第 17 回年次大会.

外山勝彦（2004）「日本法令翻訳システムの構想」『ジュリスト』2004 年 12 月 15 日号, 2-5 頁, 有斐閣.

Katsuhiko Toyama, Yasuhiro Ogawa, Kazuhiro Imai, Yoshiharu Matsuura (2006). Application of Word Alignment for Supporting Translation of Japanese Statutes into English, Legal Knowledge and Information Systems, *JURIX 2006: The 19th Annual Conference*, pp.141-150, IOS Press.

Lagoudaki, Elina (2006). Translation Memories Survey 2006: Users' perceptions around TM use, *Translating and the Computer* 28, November 2006, Imperial College London.

Pym, Anthony (2011). What technology does to translating, *Translation & Interpreting,* 3(1): 1-9.