Can Translation Software Help Legal Services Agencies Deliver Legal Information More Effectively in Foreign Languages and Plain English?

An Evaluation of Existing Translation Technology and Recommendations for Future Use

Legal Assistance of Western New York, Inc. (R)
Jeff Hogue, Esq., Technology Coordinator
and Anna Hineline

rev. April 30, 2013

We thank Maria Mindlin, Claudia Johnson, Gwen Daniels, Caroline Robinson, Sarah Winsberg, and Kari Talbott for their help and guidance. Created under TIG Project #11087.

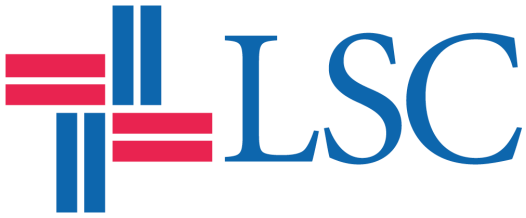


Table of Contents

Summary

This report investigates the current and possible uses of translation tools, including fully-automated machine translation and a somewhat different translation technology, translation memory. Translation memory re-uses whole phrases that an expert has previously translated. We investigate the use of machine translation, translation memory management software, and other computer-assisted translation tools for legal information in the justice community context. We also raise the apparently new question of whether translation memory libraries can be used to suggest standard replacements for legal jargon and help legal services providers produce plain language documents in plain English.

Part I. Introduction

A. Need for Foreign Language and Plain English Translation Tools in Legal Services

Translation of legal documents into other languages requires a person who is skilled at general plain language translation, fluent in both languages, and familiar with legal conventions. Consultants who can perform these translations justifiably charge non-trivial fees for such services. The small number of skilled language translators who are employed within legal aid programs and courts find themselves in great demand, and many have other duties.

As the percentage of people in the US who speak a language other than English continues to grow, legal nonprofits and courts are becoming more aware of the need to meet community langauge needs, even in areas where limited-English-proficiency populations have traditionally been small. California, Illinois, Florida, and Texas have had high linguistic diversity and high immigration trends for decades. In the past ten years however, the number of non-US-born citizens and residents has grown in states not traditionally thought of as immigrant magnets, such as Alabama, Georgia, Kentucky, and Tennessee. Because these changes in demographics are happening quickly, many legal nonprofits are not prepared to face increased demand for services and materials in other languages, and their communities and staff members may not be able to provide multi-language resources.

Moreover, there is a rising demand in the legal community generally for Plain English services and materials, particularly in the legal services context. Many legal nonprofits have devoted significant time and resources to the reduction of legal jargon to Plain English in order to better serve their clients, both native English speakers and those with limited English proficiency. Though highly beneficial for clients, the translation of documents and materials into Plain English is a time and cost burden for always-inundated legal services providers.

Providers that seek to offer a wider linguistic range of services and materials, including both foreign language and Plain English services and documents, face imperfect options. Human translation is time-intensive and relatively costly. Fully automated machine translation tools, in which a computer or software program completes translation without the input or review of a human translator, often produce rudimentary translations. Machine translation technology will someday reach the sophistication of Star Trek's universal translator, and provide instant, accurate translation into any language and dialect. Today, though, machine translation produces common errors of word-for-word translation: it fails to reflect the changed meaning of words in context. Machine-only translation results in significant translation errors that may be critical in the context of the provision of legal services.

Computer assisted translation (CAT), by contrast, holds more immediate promise in helping us lower the costs of foreign language and, perhaps, Plain English translation. Translation memory databases, one type of CAT, solve the word-for-word translation problem by remembering whole phrases of translated text. New software tools can help collaborating agencies reuse and improve translation databases that can be customized for the legal aid context and hold whole phrases that are faithful to the original document's meaning.

Our project assesses how these translation technologies can be applied to written legal language in the justice community context in order to provide cost-effective and accurate foreign language and Plain English translations.

B. Roadmap for this Report

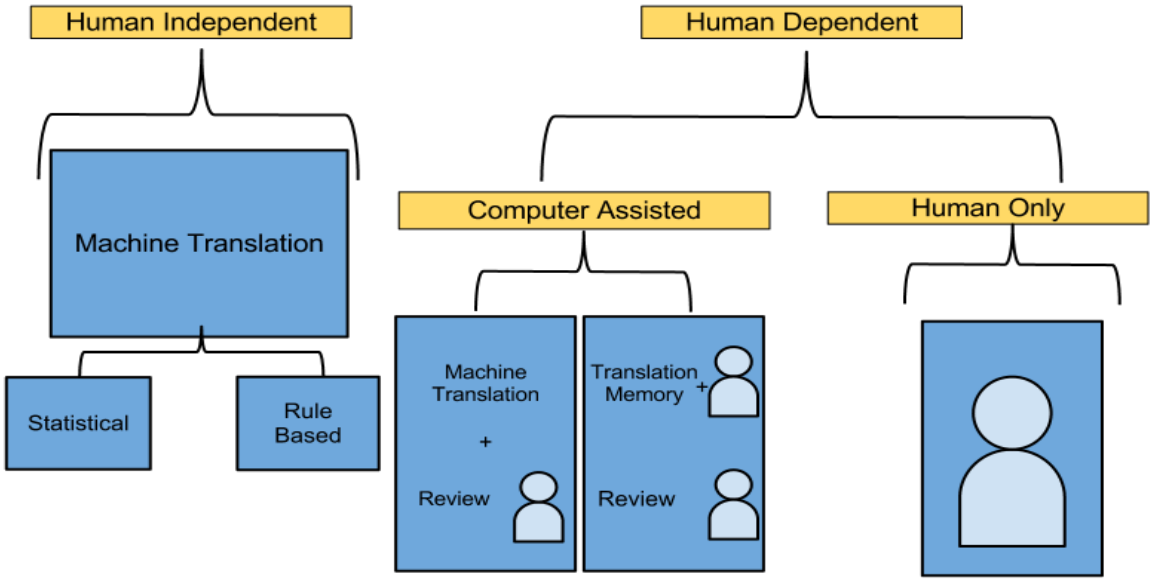
The following portion of this report (Part II) discusses various foreign language translation technologies. It gives an overview of the types of foreign language translation tools available and covers specific representative examples of each. Part III of this report details our testing and evaluation of various foreign language translation methods, and contains our findings on the strengths and weaknesses of each, with the primary criteria for evaluation being accuracy and expense. Part IV of this report covers the possible application of foreign language translation concepts and tools to Plain English translation. Part V gives our conclusions and recommendations.

Part II: Foreign Language Translation Software and Tools Currently Available

A. Overview of Types of Translation Tools

There are two main types of translation tools available: human-independent tools and human-dependent tools. Human-independent tools are those in which the translation is completed without the input or guidance of a human translator. Human-dependent tools, on the other hand, require the participation, in varying degrees, of a human translator. Computer asssisted translation, for this report, includes all methods that rely on a computer to help find and select translated words or phrases.

Organizations use each of these tools in their own unique ways that may blend the lines, and it may be more accurate to think of the use of translation tools on a continuum rather than in rigid categories. However, for ease of description and testing, we have divided the tools into categories based on the level of human participation in the process. These tools are visualized in the chart below.



1. Human-Independent Tools

Human-independent tools are machine-based, and function without the help of a human translator.[1] Machine-based, human-independent translation technology is usually called **machine translation** (MT), though other human-independent technologies may exist that we did not recognize as distinct from MT. Once data is entered into a machine translation tool, the translation is processed in its entirety by the machine. The tool will automatically translate the text, using data from multilingual dictionaries and collections of texts that have already been translated. Machine translation uses "the computer's capacity to calculate in order to analyze the structure of a statement or sentence in the source language, break it down into easily translatable elements and then create a statement with the same structure in the target language."[2]

2. Human-Dependent Tools

Human-dependent tools rely in some way on the participation of human translators in the translation process. These tools include human editing of a machine translation, the use of **translation memory** (TM) software that stores translated phrases to be selected one at a time by human translators, and pure human translation.

We broadly categorize Computer Assisted Translation (CAT) as any process that intentionally involves signficant computer and human processing of a text. A wide range of tools that fall under the category of Computer Assisted Translation. We focus on the most prevalent types of Computer Assisted Translation in this report: machine translation with human revision and translation memory applications.

a. Machine Translation With Human Editing: Cheap and Easy

Machine translation tools generate whole blocks of text that are then directly used by a client. Examples include Google Translate and Microsoft Translate, which are discussed futher below. Any agency with Internet access can experiment now with a process that begins with a machine translation. The machine translation, while far from satisfactory, saves the human translator some keystrokes when he or she sits down to perform a careful translation.

b. Translation Memory: The State of the Art for Computer Assisted Translation

[1] Of course, a team of human programmers and language experts would be required to build a worthy machine translation tool.

[2] http://translationjournal.net/journal/29computers.htm

Translation memory databases store frequently translated words and phrases for re-use in future translations. Translation memory tools are software applications used by human translators that store source segments and their corresponding translations in a translation memory database. [3]

Imagine that a group of skilled professional translators had the memories of computers. A translator could instantly recall, for example, that she had translated "If you have an order of protection in your state, New York will honor it." to: "If you have a protective order in your state, it also works in New York." A translation memory database would remember every such instance, and would also remember the carefully-crafted Spanish version of that translation. These language segments could be integrated into drafting tools for videos, brochures, and online forms. This is the promise of translation memory.

The number of translators using translation memory is growing. In 2006, Imperial College London conducted a survey of language professionals on their use of translation memory tools. Of the 874 translation professionals surveyed, 82.5 % confirmed that they use translation memory.[4]

If the justice community can gain access to a translation memory management software package at a low enough cost, and that tool is shared across multiple legal nonprofits, all of our content creators can get an instant, inexpensive 'first pass' translation into another language. This will save time and money in developing content, and will raise the nationwide standard for the documents we produce.

c. Human Translation

Some translators are so skilled and so familiar with certain kinds of translation that computer software saves them no time. As for quality, in our presentations and our discussion with language experts, no one expressed the opinion that any machine translation or Computer Assisted Translation produced better results. However, CAT should produce translations of quality equal to the skill of the human editor, and, if the acquisition costs and usability hurdles are low enough, should do so with greater efficiency and uniformity.

B. Software and Tools in Use Now

In the past twenty years, a number of software tools designed for computer-assisted translation have come into use, at varying price points and with different features. Below, we review a representative sample of these options in the context of translation for legal nonprofits.

---

[3] http://www.illinoistech.org/story.aspx/300678

[4] *Translation Memory systems: Enlightening user's perspective.* Imperial College London, November 2006

The three primary forms of Computer Assisted Translation that are currently in use are: machine translation with subsequent human editing, translation memory software, and terminology-management software.

1. Machine Translation and Human Review: Software and Tools in Use Now

Machine translation tools operate by generating results that are then directly used by a client without any additional review or revision by a human operator. Below, we discuss a sample of available machine translation tools, with the understanding that these tools used on their own are properly categorized as machine translation tools. However, while the technology behind machine translation has come a long way in recent years, it is still no replacement for a human translator, and many of the below listed machine translation tools are used in conjunction with other, human-based translation methods, bringing their use into the category of Computer-Assisted Translation, which is discussed further in Section II.B.2 below. Though this is only a sampling of the available machine translation tools, we do not expect that there are any machine translation tools currently available that eliminate the need for extensive review by a human translator.

In practice, no legal services provider is likely to rely solely on automated machine translation to translate legal information because the results often contain critical errors and poor phrasing. Machine translation alone has not been found to be effective when creating materials for the public at large.[5] However, many agencies use machine translation as a starting point, and then have a human translator edit the imprecise machine translation product. This use of machine translation in conjunction with human editing is one form of computer-assisted translation.

a. The State of the Art of Pure Machine Translation

Currently, pure machine translation produces imperfect translations, and often additional human editing and review is needed. In their article in the July 2004 edition of *Translation Journal*, Olivia Craciunescu, Constanza Gerding-Salas, and Susan Stringer-O'Keeffe note that using machine translation still requires human involvement to ensure grammatical and contextual accuracy.

> "It is important to stress that automatic translation systems are not yet capable of producing an

[5] See, How Effective is Machine Translation of Legal Information? (http://www.probono.net/library/item.317793-How_Effective_is_Machine_Translation_of_Legal_Information).

> immediately usable text, as languages are highly dependant on context and on the different denotations and connotations of words and word combinations. It is not always possible to provide full context within the text itself, so that machine translation is limited to concrete situations and is considered to be primarily a means of saving time, rather than a replacement for human activity. It requires post-editing in order to yield a quality target text." [6]

But sometime in the future, machine translation will provide acceptable instant translation into any widely-used language. On October 25th, 2012, Microsoft Research project leader Rick Rashid spoke in English on stage at a technology convention, and his words were almost instantaneously translated into Mandarin Chinese. The system didn't just generate simple word-for-word translation, but adapted his sentences to Mandarin Chinese syntax.

> In my presentation, I showed how we take the text that represents my speech and run it through translation- in this case, turning my English into Chinese in two steps. The first takes my words and finds the Chinese equivalents, and while non-trivial, this is the easy part. The second reorders the words to be appropriate for Chinese, an important step for correct translation between languages.
> - Rick Rashid, Microsoft Research[7]

We anticipate that even these advanced neural network tools will require examples gleaned from law-related conversations, legal letters, and legal pleadings, at least while the tools are being built. It is impossible to know whether these tools will be available in a few years or a decade. A less capable version of a similar tool was available in 2012, as Microsoft Translate™.

On October 31, 2012, Google launched an improved dictionary for Google Translate, and included alternate translations based on the frequency of the translation. This is a form of statistical machine translation, which requires millions of words, many tracked user selections, and is CPU-intensive. Fortunately for Google, their systems and usage give them the ability to employ statistical machine translation on a broad scale. Unfortunately, their translation "domain" is an aggregate of whatever people decide to translate; we can't instruct the tool to suggest words based on their usage in the legal domain, much less the domain of legal documents which we want to translate into plain language.

We can expect these tools to continue to improve, take on the characteristics of translation memory databases, and eventually personalize results based on how frequently a specific user selects a certain translation. This will bring the quality of pure machine translation much closer to the quality of human-dependent tools.

b. Real World Example: Illinois Legal Aid Online and Google Translate API

[6] http://translationjournal.net/journal/29computers.htm

[7] http://blogs.technet.com/b/next/archive/2012/11/08/microsoft-research-shows-a-promising-new-breakthrough-in-speech-translation-technology.aspx#.ULQgB-TO0fX

One example of the use of human editing to refine machine translation in the legal services context is the use of the Google Translate API by Illinois Legal Aid Online (ILAO). ILAO is currently using the Google Translate API to test MT with human editing under a TIG grant.

By itself, the Google Translate API tool is an example of pure machine translation. It allows built-in use of Google Translate by giving users the option of programmatically integrating Google Translate into websites and programs. The tool works by passing segments of text, as well as metadata, to Google, which Google then translates and returns translated results. The results are returned based on the specific website or program the user has integrated with Google Translate; they may be displayed on a webpage, or be sent directly to another translation program for further refinement. The API currently supports translation between over 60 languages.

ILAO's use of the tool integrates Google Translate API and a refinement program, finished by human editing. ILAO has integrated the Google Translate API into its custom-built ColdFusion-based content management system in order to create a first draft of translated legal documents. The process sends text to Google for immediate translation and then builds a new piece of content, such as a complete legal document from the returned text. In addition to the translation, their implementation preserves formatting and style and non-translatable elements such as author name, problem codes, and website tagging. ILAO employs a full-time Spanish outreach and content coordinator, who is a native Spanish speaker, who then reviews and edits the translated document. While Google Translate is free, the API is not. Google currently charges $20 per million characters translated. In the first 9 months of use, ILAO has spent about $20 using the service.

ILAO's use of the Google Translate API, including as it does review of the translation by a Spanish speaker, makes their particular use an example of MT with human editing, a subcategory of Computer Assisted Translation. Their project has so far been limited to Spanish language translations. ILAO reports preliminary findings that the Google Translate plus human review process reduces translation time costs by about 33%.

2. Translation Memory: Software and Tools in Use Now

Translation memory (TM) uses past translations to assist translators with new translations, thus improving their efficiency. Translation memory is most effective when translating multiple documents that combine repeated stock language with new content. Translation memory has long been used and appreciated by translators working for large commercial clients such as corporations. However, few in the legal aid community and in the courts have made use of this technology.

Translation memory programs store previously translated source texts and their equivalent target texts in a database and retrieve related segments during the translation of new texts. Such programs split the source text into manageable units known as "segments". As the translator works through a document, the software displays each source segment in turn and provides a previous translation for re-use, if the program finds a matching source segment in its database.

After the translation for a segment is completed, the program stores the new translation, if a new one was created, and moves onto the next segment. The translation memory, in principle, is a simple database of fields containing the source language segment, the translation of the segment, and other information about the segment's use and creation.

Some translation memory programs function as standalone environments, while others function as an add-on or macro to commercially available word-processing or other business software programs. Add-on programs allow source documents from other formats, such as desktop publishing files, spreadsheets, or HTML code, to be handled using the translation memory program.

> *Sometimes "translation memory" is a confusing term. A database that stores source and target texts is a "translation memory" as mentioned. However, sometimes a computer-assisted translation tool which imports translation memory to utilize and usually includes an editor, terminology, search function and other useful functions for translators also called "translation memory (Fig. 3)." To avoid confusion, hereinafter, we use "translation memory database" for the former, "translation memory tool" for the latter and just "Translation Memory", if it is not necessary to be distinguished. -Yasuhiro Sakine* (link)

a. Translation Memory - Software Examples

SDL Trados (http://www.trados.com/en/freelance-translators/default.asp)
Trados is generally considered the market leader in translation memory software. Its translation software is compatible with most common file types, but translation is done within SDL Trados Studio's own interface. Licenses must be purchased separately for each computer, at significant cost; however, there is no limit on the amount of content that can be translated. SDL Trados's Starter and Freelance packages are targeted at part-time or full-time freelance translators. They may be appropriate for legal nonprofits translating a very high volume of documents.

Wordfast

http://www.wordfast.com/
Wordfast is a less expensive alternative in translation memory software. It is second to SDL Trados in market share. The software is available as an add-on to Microsoft Word (Wordfast Classic) and as a stand-alone program for multiple file formats (Wordfast Pro). Limited functionality versions of both programs are available for free, and full versions are available at half or less of Trados's prices.

Of particular interest is the newest Wordfast format, Wordfast Anywhere. This free online program attempts to duplicate Wordfast Classic within a web browser. Users can incorporate machine translation through Google Translate, and can use a large set of translation memories created by other users. Groups of users can also create, store, and share their own sets of translation memories. The program will search first for an exact match for the input text, and then use "fuzzy matching" to find the most similar translation memory. If no similar translation memories are available, the program will translate the text using machine translation.

OmegaT
OmegaT is a free translation memory application written in Java. It is a stand-alone program that accepts multiple file formats as input. Like Wordfast, it uses fuzzy matching, and then machine translation, if exact match translation memories are unavailable.

3. Terminology Management Software: Software and Tools in Use Now

Terminology management software is a third type of computer-assisted translation. It works by automatically searching a terminology database for terms appearing in a document. With an accompanying software product, this functions like a sophisticated word glossary. This software also allows for translators to check if a translation memory is making the correct matches when being used in a project.[8]

It appears that the functions of terminology management software are essentially integrated into translation memory databases for translation. Nonetheless, a terminology-only translation functionality can be useful for swapping out frequently used words, especially jargon and terms of art. As such, terminology management software can be easily adapted to be utilized in plain language English translation in addition to foreign language translation. For instance, LawNY® submitted a proposal to build automatic terminology database-checking into the next version of the A2J Author software. This tool would scan interviews for complex words and suggest precise terminology swaps for frequently used words. For example, it might suggest "start" for "commence" in an interview which explains the beginning of a court action.

[8] http://en.wikipedia.org/wiki/Computer-assisted_translation

Part III: Evaluation of Existing Foreign Language Translation Software and Tools

A. Evaluation Criteria and Methodology

We evaluated existing software based on two important criteria: the quality of translations that will be produced using it, and the time and money resources required to use it effectively. It is important to note that the software must be evaluated for typical use by a legal nonprofit. Certain features, such as the ability to translate many documents at once, may not be as valuable to legal nonprofits as they are to commercial translation agencies.

1. Essential Evaluation Criteria: Quality & Resources Required

a. Criteria - Quality of Translation

High-quality translation is of utmost importance for legal nonprofits. Public informational material could be shared with thousands of people online. In private communication with clients, a small error could deprive a person of vital information and cause great harm. Lawyers must even consider whether signing and sending incorrect information may be considered malpractice in some instances. These factors create a great incentive for legal nonprofits to minimize the risk of poorly worded or misleading translation.

b. Criteria - Resources Required

Often, courts and legal nonprofits do not budget for translation in document production. Translation costs are treated as externalities, and as a result, when the costs of translation are explored in order to make the project multilingual, these costs come as an expensive surprise. Even when translation costs are known and budgeted for, as the English source material undergoes changes and revisions, the costs of translations increase, particularly if the source text is not in plain language. Most groups also don't include the cost of plain language treatment prior to translation, so they may face the cost of plain language treatment and then translation as double surprise, increasing the perceived cost even further.

In addition, most legal aid funders don't factor in translation for the projects they fund. Although some funders like DOJ do recognize translations as overhead under their grants in the administrative costs, other funders do not allow or cover language access costs as a matter of routine.

As a result, courts and legal nonprofits are eager for technological solutions that reduce the time spent by a translator, and which do not themselves greatly add to translation budgets in software and update costs. Legal nonprofits that do not have a professional translator on staff, and may not have an ongoing relationship with a professional translator, may be especially interested in CAT that will allow for higher-quality translation by fluent speakers without professional translation training.

In evaluating the expense of translation software, we consider monetary expenses including both the products' sticker price and ongoing expenses for maintenance and updates. Most importantly, we consider the products' efficiency in using translators' time, and to what extent the translation software accelerates the translation process for translators once they have mastered the software. We expect that any savings from using translation software would come primarily from reduced time spent by translators. We also consider their ease of use, and the learning curve required for translators to master them.

2. Methodology - Evaluation Process

The evaluator for translation software software is an experienced professional translator who has used a variety of translation programs. The evaluator performed translations using each translation program under consideration, and without the assistance of any program. The evaluator reported on the quality of the translations produced using each method, as well as the ease of use and amount of time spent as described above.

B. Evaluation of Results of Translation Quality Using Specific Tools

1. Google Translate (MT)

Cost: Free
Usability: Easy

A professional translator ran two samples on identity theft through Google Translate. One is written in general terms and the other is written in plain language. The results are below:
Grey highlight = awkward term or phrasing, but still understandable.

What is identity theft? Identity theft is when someone has taken identifying information about you, including your credit card or bank account numbers, Social Security number, or name and address, and has used the information to purchase items or commit a crime.	¿Qué es el robo de identidad? El robo de identidad es cuando alguien ha tomado la información que lo identifique a usted, incluyendo su número de tarjeta de crédito o de cuenta bancaria, número del Seguro Social, o su nombre y dirección, y ha utilizado dicha información para comprar artículos o cometer un delito.
What is identity theft? Identity theft is when someone takes information about you, and uses it to buy things or commit a crime. The information about you might be your: o Credit card number o Bank account number o Social Security number o Name and address.	¿Qué es el robo de identidad? El robo de identidad es cuando alguien toma la información sobre usted, y lo utiliza para comprar cosas o cometer un delito. La información sobre usted puede ser su: – Número de tarjeta de crédito – Número de cuenta bancaria – Número de Seguro Social – Nombre y dirección.

The translation was fairly accurate with sporadic words and phrases that were not the best translation, but still allowed for the material to be understandable.

The tester then ran a sample on spousal support which has language that is not as common as in the identity theft samples. The tester ran the resulting Spanish translation back through Google Translate to show the similarities and differences between the original sample and the back translation. The results are below:

Grey highlight = awkward term or phrasing, but still understandable
Yellow highlight = fatal flaw, the meaning of the English was lost

Can spousal support payments be taxed? Yes. Unlike child support, spousal support payments are taxable income to the person who receives support. They are deductible from taxable income by the person who pays support. If child support and spousal support are not given separately, then the whole amount is considered spousal support for tax purposes. Look carefully at exemptions and deductions, however, because it is still possible to not have to pay taxes on some spousal support.	¿Pueden los pagos de manutención del cónyuge pagar impuestos? Sí. A diferencia de manutención de los hijos, los pagos de manutención del cónyuge son ingresos gravables a la persona que recibe la ayuda. Ellos son deducibles de la base imponible por la persona que paga la manutención. Si manutención de los hijos y del cónyuge no se dan por separado, el importe total se considera la manutención del cónyuge a efectos fiscales. Mire cuidadosamente las exenciones y deducciones, sin embargo, porque todavía es posible para no tener que pagar impuestos sobre algún apoyo conyugal. **Back translation of above:** Can spousal support payments pay taxes? Yes. Unlike child support, spousal support payments are considered taxable income for the person receiving the help. They are deductible from the taxable base by the person paying support. If child support and spousal support are not given separately, the total sum is considered spousal support for tax purposes. Look carefully at the exemptions and deductions, however, because it is still possible to not have to pay taxes on any spousal support.

Significant parts of the spousal support sample were lost in translation and in some cases completely the meaning of the words and phrases was completely changed which could lead to a client receiving the incorrect information.

All in all, Google Translate is best used as a reference point in translation. It is free to use and is very user-friendly. That being said, it is very important to have a professional translator or a speaker of the language to review the material before it is distributed., especially if there is a lot of legal vocabulary or complex sentences.

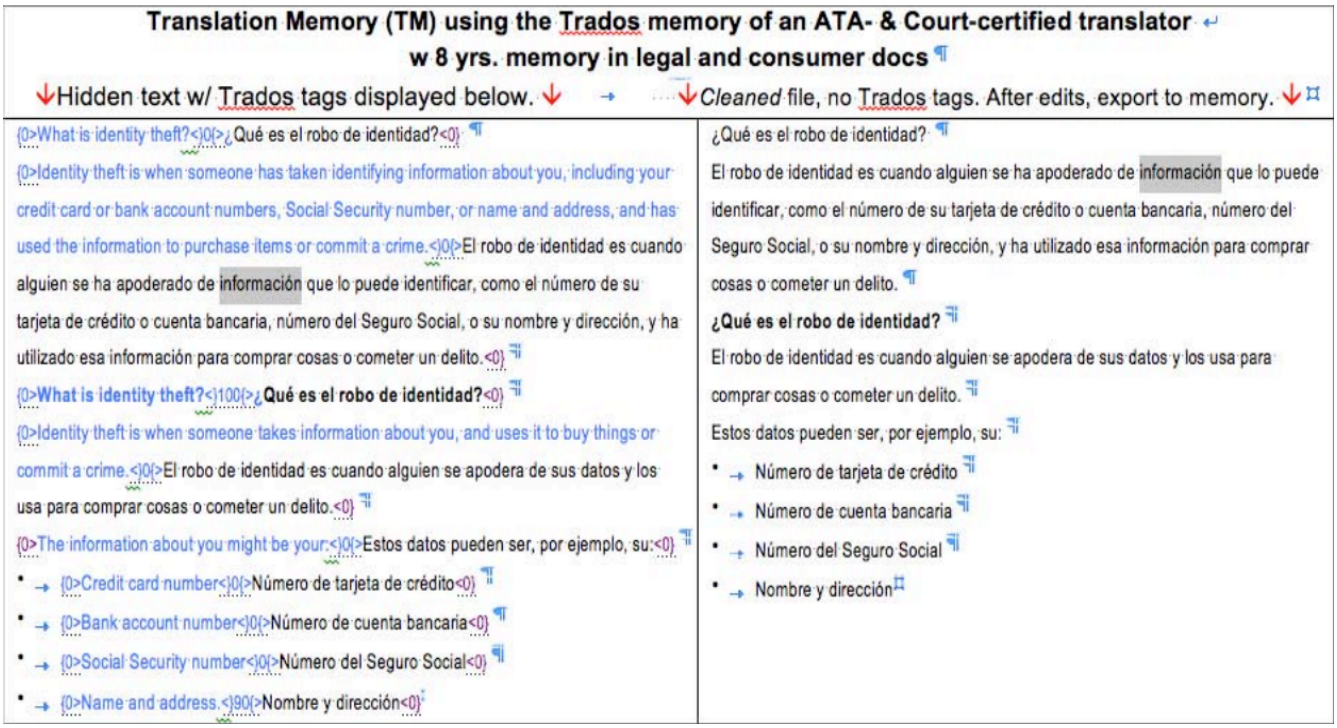
2. Trados (TM)

Price: $1085

Usability: Difficult

A professional translator ran two paragraphs, one in standard English and one in plain language through Trados. The results were as follows:

Grey highlight = awkward term or phrasing, but still understandable.



Translation Memory (TM) using the Trados memory of an ATA- & Court-certified translator w 8 yrs. memory in legal and consumer docs

Hidden text w/ Trados tags displayed below.	*Cleaned* file, no Trados tags. After edits, export to memory.
{0>What is identity theft?<}0{>¿Qué es el robo de identidad?<0}	¿Qué es el robo de identidad?
{0>Identity theft is when someone has taken identifying information about you, including your credit card or bank account numbers, Social Security number, or name and address, and has used the information to purchase items or commit a crime.<}0{>El robo de identidad es cuando alguien se ha apoderado de información que lo puede identificar, como el número de su tarjeta de crédito o cuenta bancaria, número del Seguro Social, o su nombre y dirección, y ha utilizado esa información para comprar cosas o cometer un delito.<0}	El robo de identidad es cuando alguien se ha apoderado de información que lo puede identificar, como el número de su tarjeta de crédito o cuenta bancaria, número del Seguro Social, o su nombre y dirección, y ha utilizado esa información para comprar cosas o cometer un delito.
{0>**What is identity theft?**<}100{>**¿Qué es el robo de identidad?**<0}	**¿Qué es el robo de identidad?**
{0>Identity theft is when someone takes information about you, and uses it to buy things or commit a crime.<}0{>El robo de identidad es cuando alguien se apodera de sus datos y los usa para comprar cosas o cometer un delito.<0}	El robo de identidad es cuando alguien se apodera de sus datos y los usa para comprar cosas o cometer un delito.
{0>The information about you might be your:<}0{>Estos datos pueden ser, por ejemplo, su:<0}	Estos datos pueden ser, por ejemplo, su:
• {0>Credit card number<}0{>Número de tarjeta de crédito<0}	• Número de tarjeta de crédito
• {0>Bank account number<}0{>Número de cuenta bancaria<0}	• Número de cuenta bancaria
• {0>Social Security number<}0{>Número del Seguro Social<0}	• Número del Seguro Social
• {0>Name and address.<}90{>Nombre y dirección<0}	• Nombre y dirección

This translation was more accurate than the one that was created by Google Translate, but the accuracy comes at a high price in both time and money. Translation memory tools can be very expensive and the learning curve is very steep. Even if someone is able to get trained on the tools, someone has to put in the time to build up a translation memory that is used in translating materials. If there is not a good translation in place, then the tool will be useless.

3. Omega T
Cost: Free
Usability: Difficult

Omega T was briefly tested by a non-translator. There are extensive directions on the website which have to be followed closely in order to use the software. The website even mentions that if you are the type of person who likes to learn by pressing buttons until something happens that way you want it to, then the program is not for you.

The testing was not possible to complete because the libraries that OmegaT uses for its "fuzzy matches" were no longer available for download. While it is likely possible to find a new set of libraries that could be used, the extensive training that is necessary to use OmegaT would make it difficult to adopt in programs without staff dedicated to translation.

4. Wordfast
Price: $515
Usability: Difficult

A non-translator did a brief test of the demo version of Wordfast Pro. The demo has the same capabilities as the full version. While Wordfast was more intuitive than OmegaT, it quickly became apparent that the program was useless without already having a translation memory available. While some are available to download for free online, it is hard to determine the accuracy of these memories. Due to a lack of translation memories, the tester was unable to test the translation capabilities of Wordfast.

5. Wordfast Anywhere
Price: Free
Usability: Moderate

A non-translator also did a test of Wordfast Anywhere and found it to be the most user friendly of the three programs that were tested. It is browser based which allows for collaboration. The user creates a username and password and is given a unique identification number. Documents, translation memories and glossaries can easily be shared between users. The user still would need to have access to a translation memory which can be achieved through collaboration. This program still has quite the learning curve but the guides available online appear to be very helpful.

Part IV: Imagining a Plain English Collaborative Memory Translation Tool for Legal Services

We know of no effort to use technology to partially automate plain English translations. Justice community members in legal aid organizations and the US state courts could be true pioneers by working together to craft a collaborative plain language translation tool.

A. Usefulness of a Plain English Collaborative Memory Translation Tool

Translation from one language to another is a key challenge facing courts and legal non-profits. Yet these organizations also face important challenges in making information accessible to native English speakers. Documents created by lawyers and court personnel are often full of legal jargon, with definitions and explanations not provided. The target audience may misinterpret the meaning of the information, or may simply give up and not attempt to read it.

Organizations have begun to recognize and work to fix the problem, forming a "plain language movement" within the legal profession. Many pay to have some documents and online interviews translated into plain language by an outside consultant, especially when the information will be provided to the public. Some organizations still publish jargon-filled documents without realizing they have transmitted little information to the target audience. Others have some plain-language trained writers and support their efforts with plain language legal glossaries like the California court system's Self-Help Glossary.

However, these free and in-house resources are rarely marshalled to fully meet the need for plain language legal documents produced by an organization. Futher, organizations are unlikely to expend the funds necessary to contract for paid plain English translation of every letter, brochure, form, and informational pamphlet an organization uses to communicate with the public.

B. Form of a Plain English Collaborative Translation Memory Tool

1. The Use of Foreign Language Translation Software as a Plain English CTM Tool

Translation memory may be a useful tool in helping organizations write plain language documents more quickly, and therefore more often. A plain language translation memory database would capture the revision process for documents that are edited from first-draft legal language into plain language alternatives. Both the initial legal language and the subsequent plain language "translation" would be

recorded by translation memory software. As with foreign language translation memory tools, a writer using language similar to the initial legal jargon would be presented with the plain language translation that was used previously. The California Self-Help Glossary, and other plain-language legal glossaries, could be incorporated as translation memories in order to strengthen the database.

A plain language translation memory tool could save significant time for an organization, especially if that organization frequently drafts documents that contain both stock language and targeted individual-specific language. However, a translation memory database produced by, and made available to, a large group of legal organizations would grow over time in its accuracy and effectiveness as users add plain language segments.

2. Necessary Features of a Plain English Collaborative Translation Memory Tool

We concluded that several features would be necessary to create a useful translation memory tool that would attract contributions and use by legal organizations across the United States.

Several key requirements stand out.

- The tool has to be web-based. Users need to be able to upload, download and search online from anywhere in the country. Ideally, the tool would be API-enabled, allowing users to work within their primary interfaces rather than going to an additional website.
- Adding content should be a shared endeavor: the burden of creating translations should be shared so that no group is left with the onus of adding the new translations or of updating older translations as the language becomes dated or incorrect. So, the system would need to have some crowdsourcing features that allow all who participate to share the database and to grant permission to others to modify.
- There should be some incentives to participation. Non-profits must be encouraged not only to use free existing translation memories, but also add translation memories to the database. Ideally content could be shared, and participation would require that users make new translation memories available to all, rather than creating organization-specific silos. Encouraging contribution may require staff, or some sort of administrative interface, and additional costs.
- Robustness: capacity must be large. Speed of search must be fast
- The tool should anticipate use of emerging technologies like Twitter, Facebook, video, texting, and be flexible enough to be expanded other languages.

There may be existing software the meets all or most of these criteria. Tools like Wordfast Anywhere are web-based designed for large-group collaboration in creating comprehensive translation memory

databases. If no existing software could accommodate a crowd-sourced plain-language translation memory database, we should consider what would be required to build one.

The amount of time and effort that needs to be put into developing and maintaining a high quality glossary and translation memory is non-trivial. Because of the collaboration feature, a tool similar to Wordfast Anywhere would be a good starting point. Existing materials such as the Sacramento Superior Court Legal Glossary and plain language libraries that could serve as building blocks for a plain language translation memory.

3. Statistical Machine Translation

Statisitical machine translation analyzes large numbers of documents to make smarter guesses about appropriate translation. While this is part of the development of machine translation, our community has a specialized vocabulary, so we can't count on large public machine translation tools to translate legal jargon.

In the near future, however, statistical machine translation, combined with a large number of plain English documents provided by justice community partners, could acheive results that are as useful as the plain English translation memory tool we are envisioning. In other words, we could ignore translation memory, and seek out statistical machine translation technology that analyzes completed works to dramatically improve the Machine (only) Translation tools.

With a large enough corpora of plain English legal documents, statistical machine translation could meet our plain English translation needs. We would submit hundreds (perhaps thousands) of plain language examples to be analyzed by a tool that produces statistical machine translation.

This SMT approach would require:

- a machine translation software platform
- hardware to peform the statistical analysis
- contributions of hundreds or thousands of plain language legal documents for the system to glean statistics from

An exciting potential option is to build a layer on top of Google Translate or Microsoft Translate a tool that picks from the returned list a word or phrase that appears with highest frequency in our plain English language statistically-scored database of documents.

Part V: Conclusions and Recommendations

A. Machine Translation with Human Editing for Foreign Languages

We can expect machine translation software tools to continue to improve, take on the characteristics of translation memory databases, and eventually personalize results based on how frequently a specific user selects a certain translation. This will bring the quality of pure machine translation much closer to the quality of human-dependent tools.

They are going to catch up, and eliminate need for translation memory by subsuming translation memory methods into their algorithms by watching lots of users select translation segments. However, we do not expect the development of MT to eliminate the need for additional tools in the next few years. We predict it will be at least a decade before a human can perform merely trivial review of automated legal information translations.

B. Translation Memory for Foreign Language Translation

A variety of factors need to be taken into account if the justice community is going to adopt translation memory as a primary source of translation. These translation programs are very costly and take a significant amount of time and effort to learn. The professional translator who completed the quality testing stated "A legal service agency wishing to use a TM tool, would need to (1) get or buy a TM app (2) learn to use the app – which is a bit painful, (3) build robust, high quality memory based on previous translations related to the topics of future translations or by aligning reliable SP-EN documents and glossary files from which to build memory."

LSC should communicate with the National Center for State Courts to discuss creating a pilot to experiment with a select group of courts and legal aid organizations. Private foundations should be brought on board also, for seed money. Organizations such as ILAO and PBN could be brought in to offer their LEP and technical expertize. This is doable, is timely, and won't happen without collaboration among multiple partners.

C. Human Editing for Foreign Language and Plain English Translation

We have found that even when qualified translators who speak the target language create artful translations, other native speakers may disagree with the translation. All translation is subject to style and complexity preferences. As an analogy, consider the sometimes heated disputes lawyers have when trying to use plain English language in legal documents. In our experience, no lawyer-editor is

ever fully satisfied with another's work. The aim of computer assisted translation can be to save time and money and produce excellent translations, but cannot be to produce translations that are unarguably ideal.

Human translators are particularly useful in employing plain language techniques not captured by translating words alone. For instance, chunking, use of white space, editing harshly for brevity, and choosing to add helpful graphics are all techniques that computers are unlikely to suggest or employ effectively in the near future.

D. Plain English Colaborative Translation Tool

1. Translation Memory Possibilities

The amount of time and effort that needs to be put into developing and maintaining a high quality glossary and translation memory is non-trivial. We recommend that the Legal Services Corporation convene a group of leaders from legal services providers, plain language experts, and court leaders to adopt or discard this approach.

2. Statistical Machine Translation Possibilities

With a large enough corpora of plain English legal documents, statistical machine translation could meet our plain English translation needs. This would not elminate human review, but could greatly speed the process of plain language translation.

We recommend that LSC convene a focus group of technologically-skilled community members to discuss the possibilities for selecting a machine translation software platform, hosting hardware to peform the statistical analysis, and gleaning contributions of hundreds or thousands of plain language legal documents from justice community partners.

The group should also conceptualize the requirements of building a layer on top of Google Translate or Microsoft Translate that picks from the returned suggestion list the the plain English translation most preferred by our legal community, as determined by statistical analysis of the copora of plain English legal documents we have submitted into the database.

Appendix A: Papers and Articles on Computer Assisted Translation

Phrase-based Memory-based Machine Translation
Maarten van Gompel
HAIT Master Thesis
Tilburg University, The Netherlands

Extending Memory-Based Machine Translation to Phrases (Presentation)
Maarten van Gompel, Antal van den Bosch, Peter Berck

Integrating Machine Translation with Translation Memory: A Practical Approach
Panagiotis Kanavos and Dimitrios Kartsaklis

Translation Skill-sets in a Machine-translation Age
Anthony Pym
Universitat Rovira i Virgili, Tarragona, Spain

The Benefits of Translation Memory and its Misconceptions
Precision Language and Graphics, Inc.
Blog post briefly highlighting benefits and misconceptions surrounding Translation Memory

Development of Translation Memory Database System or Law Translation
Yasuhiro Sekine, with Yasuhiro Ogawa, Katsuhiko Toyama, Yoshiharu Matsuura

http://www.opentag.com/tm.htm

Translation Memory Goes Open Source: (Global By Design)
http://www.globalbydesign.com/2010/07/08/translation-memory-goes-open-source-with-open-tm2/

What is Machine Translation? (Systran)
http://www.systransoft.com/systran/corporate-profile/translation-technology/what-is-machine-translation