Collaborative Translation Environment for Large Discovery Projects

Jianqing Wu, Ph.D.

Summary

A server-based translation environment[1] and related translation protocol are now available for improving translation performance for cases where a large number of documents are generated in a source language context but their controversies are adjudicated in a different language context. The technologies, which are better alternative to Translation Memory System in the legal field, are intended to solve five oldest and toughest problems: translation task-management, terminology inconsistency, common, inherent, and distinctive errors, language contextual shift, and incorporation of a foreign language context in documents created by second language speakers. The technologies are suitable for any kinds of cases involving any languages, are available for local deployment, and offer the highest document security. If the translation protocol is used by well trained and motivated document reviewers in a collaborative and harmonic setting, it can reduce unnecessary translations, improve translation accuracy and readability, minimize the needs for amendments, contain translation costs, and help the client significantly improve its litigation position.

A. Problems in Foreign Language Review and Translations

It is well known that foreign clients have obviously lower chances to win their cases in a litigation forum. It is often believed that hearing officers are bias against foreign litigants. While protectionism and local bias often play a role in many cases, the real culprits are documents and translations. It is the misleading, confusing, and often incomprehensible documents and translations that give hearing bodies ample excuses to make any decisions.

Difficulties to process foreign documents are in part due to corporate documents. Corporate documents normally contain (1) information on a large number of products, services, projects, and processes, (2) strange codes or causal names of products, services, and materials, (3) a large number of players such as employees, customers, attorneys and consultants, and other parties, (4) technical subjects of varying complexities, (5) jargon,

---

1. Based upon the key features disclosed in one granted U.S. Patent Application and three pending U.S. Patent Applications.

abbreviations, and acronyms, (6) assumptions known by those involved in underlying transactions, (7) incomplete person names, place names, and discussion topics understandable to only those in the discussion group, (8) all kinds of historical usages and random errors, (9) patent and trade secrets protected by passwords, and (10) texts in one or more foreign languages. In addition, corporate documents contain a large number of duplicates, which may be from hundreds to thousands of copies. Other factors making the problems worse include a large number of file formats, encoding problems of foreign language in documents, and password protection of documents.

Two common characteristics, cross-document verbal context and incorporation of external facts, further complicate the review and translation of documents. Verbal context in corporate documents is different from the counterpart concept in linguistics. This can be shown in a contract dispute case. Relevant documents in a contract dispute may cover a series of subjects from initial marketing effort, initial contact, preparation of bidding documents, tendering of bids, formation of contract, post-bidding activities, delivery of products, making payment, occurrence of natural disaster, request for change of price, breach of contract, termination of contract, filing a lawsuit, retaining law firms, settlement discussion, drafting settlement agreement, to signing settlement agreement. It is a standard practice that corporate staff does not need to mention everything that has happened before. The verbal context in most documents is deficient in themselves, but may be found in other related or precedent documents. Many documents do not provide all important facts. The documents may also freely incorporate facts commonly known in the corporation and rely on facts that the intended readers know. The documents never include background information. In the eyes of document reviewers, documents contain unexplained terms, implied assumptions, mysterious persons, unknown product names, unknown serial number, unknown locations, unknown times, unknown purposes, and unknown anything and everything.

The disruption of language context caused by current document assignment in the current document review model makes it even more difficult to ascertain "transaction context." At the current state of art, document review is nothing more than classifying documents by the best guess. Accurate understanding of documents is impossible, even though fair accuracy can be achieved in coding documents. There is no way to accurately translate documents out of context. Due to all of the above factors, most documents can only be translated verbatim. It is inevitable to result in translations that do not reflect original meanings or even become

"accurate" nonsense (a perfect translation that completely departs from the original meaning).

The nature of tasks in litigation discourages the use of new technologies. Translation memory system, which can store prior translated materials for subsequent use, cannot be used due to its high deployment costs, risk of leaking confidential and sensitive information, and potential recycled errors, a well known problem. When translation is performed by off-site vendors, translations are conducted in a context-deficient environment, inevitably resulting in massive distortions and fatal mistakes. As a result, translation is still same as it was since it is first available. It has not be able to avail all kinds of information technologies.

B. Consideration of Large Litigation Models

To understand translation problems in discovery, one should examine two representation models. The most efficient representation model is a single super lawyer representation model ("SSL") by which a single lawyer does all legal services. In this model, the lawyer (1) knows all laws; (2) makes all strategic decisions and makes all necessary changes; (3) reviews all documents and thus understands all case facts; (4) knows all elementary facts and their significance; (5) knows all legal issues; (6) does not have any internal communication problem; and (7) has no divided royalties, no conflicts of interest, and has little concern with his own liability. Therefore, the super lawyer can deliver superb legal services at the highest productivity.

The information expansion has forced the whole industry to use Multiple Players Tasks-divided Representation Model ("MPTD") by which one or more law firms or a group of lawyers represent a client, with necessary supports from document reviewers, consultants, and data-hosting companies. This model has the following characteristics: (1) strategic decisions are made by lawyers with different expertise; (2) case knowledge is divided among many lawyers and many reviewers; (3) everything is done by N-folds duplicate effort; (4) there are serious communication problems among lawyers, document reviewers, consultants, and supporting staff; (5) documents concerning a single transaction, event, or fact are assigned to several reviewers; (6) due to disruption of the cross-document context, each document is reviewed on its face and its significance is determined on the face and attachments by the "best call;" (7) conflict-of-interest rules have facilitated the hiring and employment of inexperienced legal staff and inexperienced reviewers; (8) the lack of experience and special knowledge of review process by the management can dramatically increase review time, greatly reduce review

quality, and increase the risk of exposure; (9) intensive competition among staffing agencies forces them to select review candidates by speed, and reduce reviewers' qualifications to the minimum; (10) there is an obvious distrusted relationship among all participants; (11) clients fail to see the importance of document review but believe that it takes only a great legal argument to win; (12) document reviewers lack super lawyer's passion to deliver the best performance due to a total lack of financial incentive; and (13) in foreign language cases, the review industry lacks necessary tools, selects misfit translators, uses improper translation methods, and often provides context-deficient and tools-stricken translation environments. Millions of major and fine actions and decisions, that may take place without sufficient collaboration, may be inconsistent and even in conflict with a final disposition that the client seeks. No matter how big a case is, the final disposition is only one. To achieve a best result, all actions and decisions must be made consistent with the final goal.

C. Translation Sever Environment

Accurate translation of documents requires three sources of knowledge: general knowledge in both languages, background knowledge of underlying technical art, and specific knowledge of the case. The type and amount of language-related general knowledge depend upon who creates original documents and what are the source and target languages. The type and amount of background knowledge depend upon the services, products and transactions of the client. The type and amount of specific case knowledge depend upon the nature of specific case.

A translation environment may be created by configuring a discovery information management system ("review-assisting system"). FIG. 1 shows the actual screenshot for a configuration that can be used for all kinds of cases. The system can be changed from a language review system in FIG. 1 into an antitrust review system in FIG. 2 instantly. This system has been developed as early as 2009 and it has been tested by auto data-feeding programs. From the result, it can easily handle tens to hundreds of reviewers, depending upon host machine's computing power.

The translation environment has the ability to manage a large number of databases adaptable for all kinds of cases, and includes data retrieval tools for showing data records from database tables, search tools for searching data records from the database tables, and data entry tools for entering data records to any database tables. It also includes tools for conducting

global search, interactive search, editing whole table, looking up data (data presented in certain ways), validating data records, searching names, processing name lists, and searching personal data on the server. The translation environment allows project managers to change the table structures of all data tables instantly and reliably.
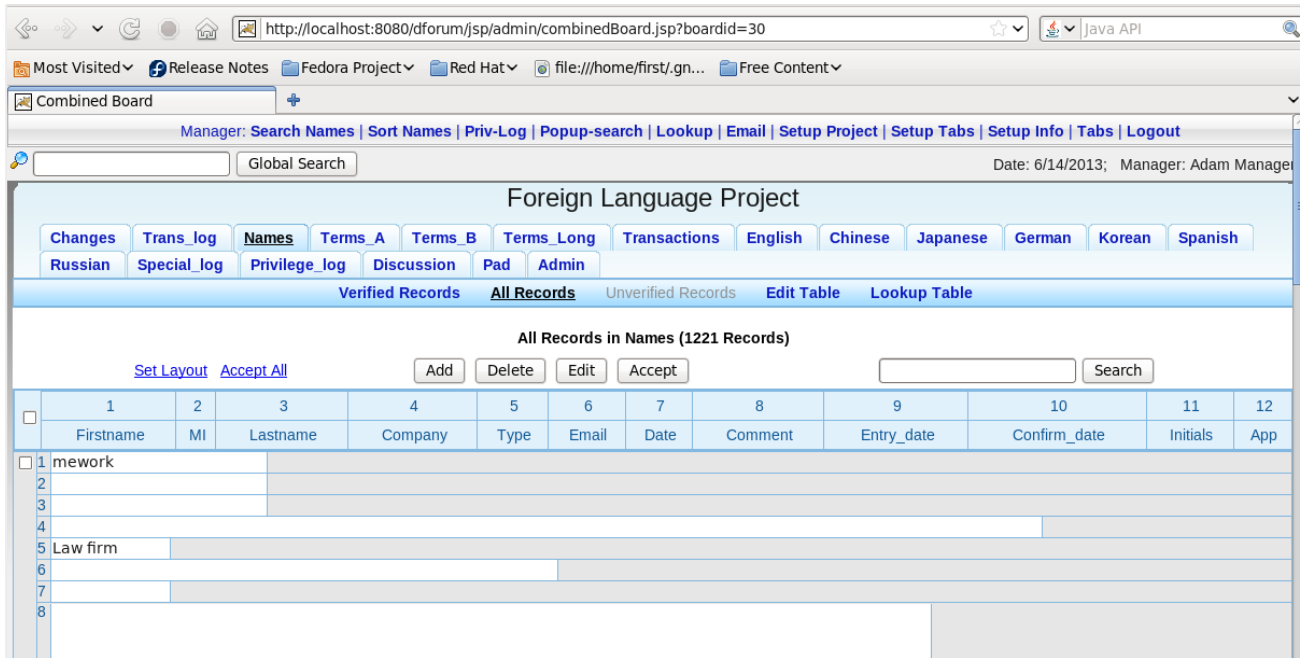

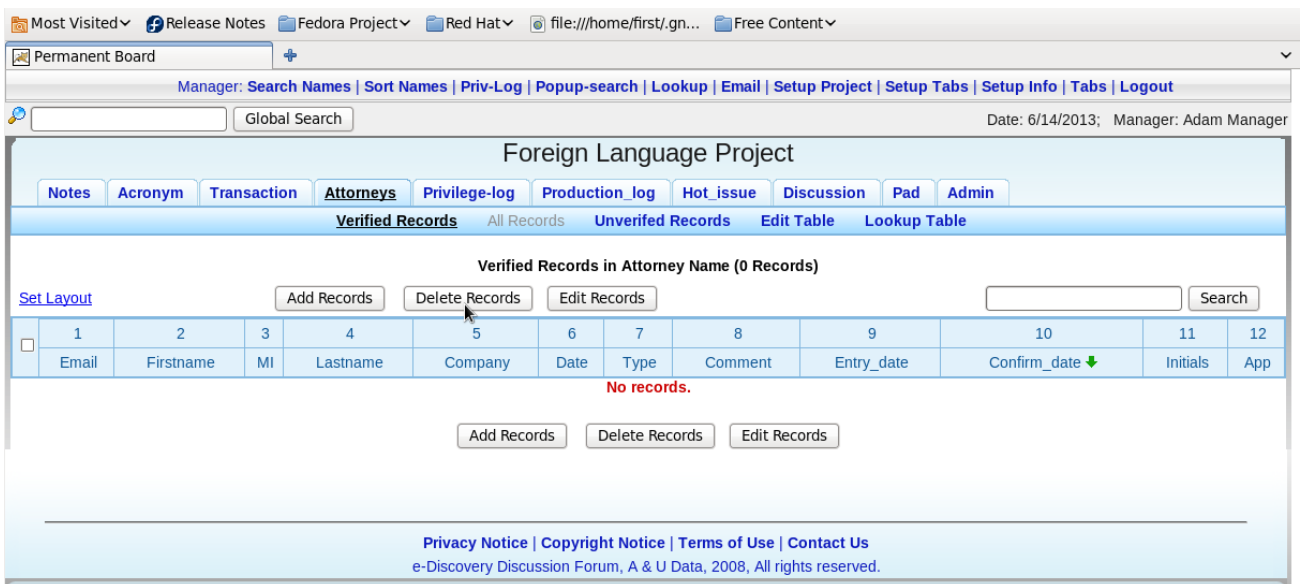FIG. 1 Discovery Information Management System for Several Foreign Languages.

FIG. 2. Discovery Information Management System for Simple Merger Case

The central concept of the translation environment is to "unify" all attorneys as a super attorney and all document reviewers, including translators, as a super reviewer. The method helps the reviewers reconstruct the relationship of segregated documents and allow the reviewers to review documents in a virtual language context.

The related translation protocol facilitates collaboration among all parties and players. It allows all players including litigation attorneys, document processors, project managers, and review attorneys to update their instructions, work products, and anything real time so that other players know all changes and updates, and thus eliminating the room for misunderstanding. The sharing of critical review information in real time helps reviewers, including translators, reduce errors.

The translation protocol helps reviewers share work products during the entire review process. By this process, the reviewers can share their work products, reuse elementary work pieces, and avoid conducting duplicate effort of conducting research, analysis, and fact-finding. As a result, the reviewers can avoid working on duplicate tasks and use the saved time to hunt for important information.

The translation protocol allows managing staff to instantaneously change their review standards, construction rules, coding rules, methods of treating documents, and factual entries. By using this system, the whole review team can reduce the amount of time for fixes, reworks, and corrective review to the minimum theoretically possible.

The translation protocol is designed so that conflicting instructions between litigation attorneys can be minimized; different interpretations of same rules, instructions and guidelines by different reviewers can be minimized; conflicting and inconsistent treatments of documents can be minimized; coding errors caused by different understandings of the same rules and instructions as a result of reviewers' different experience can be minimized; and documents-selection errors can be minimized.

The translation protocol is for maintaining a dynamic translation task tracking table, which helps all litigation attorneys and reviewers instantly determine if a similar or same document has been translated, and if yes, the attorneys and reviewers can check the original documents and translations in real time before a decision is made as to whether to translate a particular document.

The translation protocol is also a process for reducing term multiplicity for certain

matters such as personal names and titles, company names, division identities, times and duration, location names, and any of other terms that are important to the understanding and performance of documentary evidence.

The translation protocol is for minimizing the effects of a change in language context on hearing officers' ability to understand translations. Foreign language documents created in a foreign context are intertwined with a huge number of non-text contextual elements. Verbatim translations, even at the highest accuracy, are "incomprehensible" to hearing officers. The protocol is for minimizing the effects of a shift of language context on the comprehension of the hearing officers.

The translation protocol is for rephrasing the texts created by second language speakers in a target language in a foreign context. The method includes a server-implemented method for correcting obvious errors, common errors, various forms of quoted foreign terms, and a scheme for providing annotations to compensate for the effects of a contextual shift.

D. Five Oldest Problems That Make Translations Incomprehensible

The translation environment is intended to solve five oldest and toughest problems: (1) translation task management, (2) common, inherent and distinctive translation errors, (3) translation terminology inconsistency, (4) meaning change due to a shift in language context, and (5) foreign "context" in documents created by second language speakers. It can also be a best tool for handling troublesome and critical documents.

1. Translation Task Tracking Method

One important thing in the management of work flow is to determine which documents have been translated and which documents should be translated. In the complex presentation model, there are several litigation attorneys and tens to hundreds of reviewers and/or translators, it is not an easy task to manage work flow. Two common factors make this seemingly straightforward task very difficult. One of the factors is that litigation attorneys are unable to read documents in the source language, and, due to "translation term multiplicity" and "structure multiplicity," the attorneys are unable to determine whether two original documents are similar or identical by looking at their translations. An original document may end up with different translations even if the translations are product of a same translator. When different attorneys and a group of reviewers and/or translators work at different locations, this task is practically impossible. Law firms now use some simple tools to keep track of translations, it is a

trial-and-error process. This tracking process can consume a great deal of valuable time. It is inevitable that many similar or identical documents have been translated many times, while some important documents are lost in the forest.

Terminological, structural and stylistic multiplicities complicate the task of tracking translations by those who cannot read native documents. Tracking translations become extremely difficult under the multiple-player task-divided representation model. The degrees of difficulties is exponentially related to the number of documents and the number of participants involved. In some cases, litigation attorneys may pick up documents for translation on the basis of reviewers' comments, but are unable to tell whether particular documents have been translated. When two or more litigation attorneys work in the same case, their inabilities to evaluate original documents make the situation even worse. In other cases, the task to designate documents for translation is delegated to document reviewers, but the reviewers do not have a suitable tool to check out all translated documents. Each reviewer makes determination on the basis of what the reviewer knows. Different reviewers may pick up the same or similar documents for translation. When a review project is going on for months or years, one reviewer may translate many duplicate and similar documents. This problem alone can easily double or triple total translation time.

A good solution is to use a translation-tracking table of the Discovery Information Management System. It allows all reviewers/translators to track translations by digital properties, document description, author and recipient information, and control number. It allows the reviewers to access original documents instantly and track translations by directing looking at the original documents in the most convenient environment. Due to the way of implementation, they can look at the same database any time, regardless of where they work.

2. Inherent, Common, and Distinctive Translation Errors

Between any two languages, a translation can have a large number of potential inherent, common, and distinctive errors and inaccuracies. The sources of errors include (1) misuse or forced use of words or phrases due to lack of identical or similar words and phrases; (2) inaccuracies introduced by the closest words; (3) impermissible words combination in the target language while they are acceptable in the source language; (4) meaning change or distortions under the target culture; (5) inversion of social values as a result of a shift of language context; (6) certain distinctive problems caused by writing styles in a culture; (7) historical word usages

inconsistent with common meanings; (7) errors and omissions in original documents; (8) missing information in the original text but required in the target language due to different grammar rules; (9) errors from failure to correctly use active and positive voice; (10) impossibility to preserve the same ambiguity in an original document; (11) uncertain meanings caused by word order; (12) meaning of a word combination being different from actual meaning; (13) loss of all hidden and secondary meanings in translations; (4) unknown special meanings known to those in the communication cycle; (15) meanings influenced by undisclosed author's personal knowledge; and (16) undisclosed external facts, context information, and assumptions.

Any translation contains common problems including imperfections, approximations, unavoidable distortions, and even human errors. Those things may have sufficiently serious impact on outcome of litigation. Those errors can be classified into four types: (1) common or inherent errors for any given source and target languages, (2) common or inherent errors in a given case, (3) distinctive and obvious errors which can be easily recognized, and (4) random and unpredictable errors which cannot be easily recognized. Very little can be done from the technical point of view to detect random and unpredictable errors.

Solution to common or inherent errors for any given source and target language: The first type of errors and inaccuracies can be reduced by loading and providing the database of the common errors which are well known for given two languages. This database may include several hundreds to thousands records in the form of source words, potential target equivalents, and usage explanations. It is as simple as searching original text using the database records and highlighting all found words in the original text with a "reference table" showing explanations of all potential errors and inaccuracies. This explanation helps the reviewers immediately know how to correctly translate those words and phrases. A powerful database table can help the reviewers avoid errors and inaccuracies. The reviewers can also add records to the database for this type of words and phrases as long as they are not distinctively associated with the case, but may be repeated in other situations and other cases.

Solution to common or inherent errors for a give case: It is impossible to know and predict any of such errors in a given case. Therefore, the solution relies upon the collective intelligence of all reviewers. This type of errors may include company-wide historical misuses, matter related to personal knowledge, and distinctive personal misuse. The words and phrases

giving rise to this type of errors can be entered into a distinctive table or shared table in a specific type. Whenever a reviewer or translator identifies an error in this nature, a record including an explanation is entered into the data table. Such a newly entered record is immediately available as guidance: any document text is searched against the database, and found words are highlighted with a "reference table" showing explanations for all highlighted words. With time passing, more and more errors are identified and the database contains more and more such terms. If it is necessary, a reconciling review may be conducted. Potential documents for a reconciling review can be identified by finding documents that contain those keys and have been reviewed before the creation of related data records.

3. Translation Term Multiplicity and Structural Multiplicity

Different documents may contain the same term to mean different things. This phenomenon is referred to as "term conflict." In contrast, a thing, a concept, a person, or a location may be referred to by different terms in different documents or in different sentences in the same document. This phenomenon is referred to as "term multiplicity" in this writing. When term multiplicity exists in large numbers in translations, it can confuse readers. Both term conflicts and term multiplicity in translations becomes serious problems in large discovery projects.

Each language allows writers to freely use different sentences, adjectives and verbs. This is so called ways of expression. A careful observation leads to a finding that expression diversity is generally not allowed in person identities, entity identities, company divisions,

product names, trademarks, logo, important events, important legal actions, key technologies, and important milestones. Those terms are used consistently in corporate documents and in the public. Occasionally, some terms may have a limited number of variants and abbreviations,

but they are recognizable by corporate staff and the public.

Corporate executives normally write their names in certain ways: some like to use full names; some, partial names; and others, abbreviated or short names. Their names may have several potential translation choices: sound translation of full name, sound translation of preferred name, correspondent foreign name, and correspondent preferred name. Some executives may use arbitrary foreign names, which are not related to their native names. Most corporate official titles also have many different ways of translation. A Chinese corporate title may have five or more ways of translations. Many entity names may comprise nation name, city name, or locality name, and one or more ordinary words. Nation's names, city names, and locality names all have different translation choices. Ordinary words include nouns, adjectives, verbs, and action words, which have even more translation choices or synonyms in any language. When ordinary words appear in company or entity names, they increase the ways of translations. The number of translations is directly proportional to the number of ordinary words in entity names. A corporate structure may also have different ways of translation. For example, "sales department," "North America region," and "field office" may be translated as "sales division", "North American continent" and "representative office."

Each of the words in a typical email header or an address block may have two to five translation choices, which could result in a large number of combinations. It should be noted that two address blocks with two-word differences could be considered as two different persons in many cultures. When one address block is different from another by three to ten words, there is no way for the readers to know whether they point to the same person to different persons. In determining whether two address blocks are for same person, native speakers can "see" certain signs and can make a good guess from the signs. However, the person reading the translations cannot see any signs, and thus cannot understand whether the two address blocks point to same or different persons. Documents with serious term inconsistency are unfit for merit trial.

Three additional factors make it even harder to identify persons and entities in translations. When thousands of documents are translated without any constraints, translations may contain all potential f translation variants, which may be in hundreds or more. Even worse, many companies have two or more employees having similar names and similar titles, and their address blocks also have many ways of translation. It is entirely possible that the identities of

an identical executive in different translations may appear to be so different that they would be considered as different persons; whereas the identities of two different executives may look so close that they are considered as same. The unrestrained freedom in translating identity terms makes translations incomprehensible and places a huge burden on the trier of facts. Moreover, many sentences may be translated with different wording structures, making the problem even worse.

The last factor is translation styles of all translators. Translations in a typical production pool exhibit a broad range of characteristics, as a result of different writing styles, technical knowledge levels, understandings of litigation, and, most importantly, translation philosophies. Translations by computer algorithms can be anything from complete accuracy to total errors; translations by those who do not understand "the nature of litigation" may show varying distortions, inaccuracies, and even errors; translations by those who care about original meanings may include additional annotations and repeated amendments; translations by a vendor in a context-deficient environment may result in rewriting of "accurate" stories. Translators who understand both two language contexts and litigation nature may feel helpless due to the difficulties to resolve conflicts caused by two language contexts. Translators who know only the target language, can translate anything because they do not see the foreign context and thus do not see the conflicts caused by two language contexts. A single translation with massive stylistic, phraseological, and terminological inconsistencies are hard to decipher in one or more aspects, but a collection of such translations make the problem worse.

Term multiplicity may affect client's interest in several ways. First, in government investigation cases where the client's documentary evidence is confusing and misleading, the government may demand additional documents and take additional depositions. This increases litigation costs. Second, confusing and misleading translations may be responsible for protracted and undesirable litigation. If a client can produce clear evidence to address the government's concern, the case may be deposed of much faster. If documentary evidence is impossible to understand, the parties can only drag it because there is no way to hear it on merit. Finally, when time is run out, the parties may have to settle the case, not on the basis of merit, but purely due to the need to get-over-with-it. The end result is that the should-be-winner may be a loser, or an innocent party may have to pay a fine. In private civil cases, both sides may suffer from protracted litigation. In governmental investigation cases, it is hard to imagine any scenario where producing incomprehensible documents can help the client. A single

instance of discussing a questionable transaction in a large number of documents may "become" many illegal transactions done by many employees due to translation term multiplicity. This is why a vast number of foreign cases do not have a day in court.

4. Incomprehensible Translations As a Result of a Shift in Language Context

It is generally agreed that language context is the foundation of any language. When the context is changed, the meanings of words and phrases are changed. While experts generally consider language context as the most important thing, none has ever proposed any solution to this oldest problem which always exist in cross-border litigation. Verbatim translations can accurately preserve original meanings only when the intended readers understand the original language context. In cross-border cases, this critical assumption always fails in all cases. The intended readers are hearing officers and investigators who know little or nothing about the original language context. In their eyes, some translations gain different meanings, some become confusing, and others are incomprehensible.

"Language context" or "context" affecting interpretations includes all text-related concepts ever discussed in linguistics, plus non-text components which may carry supplemental meanings for a subject under discussion or affect the interpretations of words and phrases in the text. The non-text contextual elements include social concepts, local culture, name convention, business practice, science practice, farming practice, human interactions, national conditions, geographic locations, local weather conditions, government organizations, legal environment, common knowledge, special knowledge, author's subjective belief, author's personal knowledge, external facts, historical events, famous persons, common mistakes, historical usages, inhered usage errors, omitted assumptions, and distinctive personal usages. Any of the components may comprise a gigantic number of fine elements.

Language context affects original meanings by providing supplemental information. Anything such as human names, company names, and locations carry a great deal of information. For example, when an author writes email in foreign names and foreign company names, the foreign names and company names carry some information about the person and the company's business. Language context may affect the scopes of words and phrases by default. If the context in some way defines the scope of a word, the change of context directly affects the scope of the word. The number of words and phrases affected by language context is extremely large, and some examples are medical products, human being, marriage, and

residency. Language context also influences the interpretations of texts. If an original text is placed into a different context, the text gains different and often confusing meanings. If the context affects the interpretation of the whole text, a change in the context may make the whole text incomprehensible. A tax mitigation strategy makes sense only under the intended taxation system. When the context the original document relies upon is "dropped out," the document loses all intended "supplemental information" so that "it makes no sense" or "it does not add up."

Even if original documents are accurately translated, the resulted translations still have serious impact on the adjudication of cases because the translations may become incomprehensible to hearing officers as a result of a shift of language context. After a document is accurately translated, it would still inhere all elements of the foreign context. The translation would be incomprehensible to the readers who know only the context in the target language. The readers will have no clue as to the players, locations, unique geographic conditions, business environment, legal environment, cultural concepts, and unique expressions. The hearing officers thus read words and phrases that either lack foundations or are attached with wrong or distorted foundations. Many things "do not add up." This problem alone, without any additional errors and term multiplicity, would easily alter litigation course. Deciding a case is applying law to facts. When hearing officers are unable to understand facts correctly, there is no way to reach a right result. This problem together with common translation errors and translation multiplicity makes most foreign cases unfit for trial on merit. The extremely high settlement records in foreign language cases fully support this point.

When cases originated in one culture is heard in another culture, their results often surprise litigants. However, if those cases are carefully examined from the target culture, many of the results are not surprising at all. The cases perceived in the foreign context are different from the same cases perceived in the forum context. The effects of context shift include: (1) the foreign context is lost in front of the hearing officers knowing only the forum context; (2) all information provided in cross-document context may become less obvious; (3) words and phrases may become incomprehensible due to a change in context; (4) the meanings of words and phrases are changed; (5) the scopes of collection nouns are enlarged or narrowed; (6) some original word combinations become unacceptable in the target language; (7) author's feelings and attitudes are changed; (8) cultural meanings of words and phrases are stripped and different meanings are forced into; (9) the social values of identical conducts are changed;

(10) the degrees of modifiers are changed; (11) external facts and unstated assumptions disappear without any vestige; (12) improper usages become more confusing; and (13) original errors and omissions lose their vestiges. Such documents are hard to understand due to all kinds of mishaps.

Any of those factors may decisively change litigation outcome. In a contract dispute where a party is sued for violating a non-competing agreement prohibiting the sales of "office products," the outcome would depend upon whether the sold items fall within the scope of office products. If hearing officers take as granted the Unites States as context, the company should win. However, the company may lose the case in a foreign forum if the trier of facts does not regard the disputed items as office products. The only way to prevent an unintended outcome is use of extensive annotations in the translation or provision of expert testimonies on the scope of the term.

When an employment agreement is drafted in the United States, it may lose a lot of the claws when it is litigated in a foreign culture. Employment contract often contains words and phrases that are sensitive in any culture. Hiring employees in Brazil is easy, but termination of employees without any liability is difficult. An employment contract may be unable to protect Brazil employers as it could in the United States. When a party takes a position that is strongly against a cultural value in a foreign forum, the chance to win is slim.

Any conduct relating social value in an adjudicatory forum may invite a hostile or sympathetic reactions from the trier of facts; the stripping of verbal context and non-verbal context may impair perceived facts, as understood by hearing officers; improper usages, errors and omissions may distort critical facts; even sentence structures and word orders could affect litigation outcome if the central issue is construction of law and application of law to critical documentary evidence.

Solution to the context shift problem. If a dispute is revolving around the scopes of words and phrases, provision of expert opinions is the most proper option. However, in a large case, it is unrealistic to have experts to review all documents and write opinions for them. A viable solution is to produce most accurate translations with annotations. If conduct in documents has an unfavorable effect on the client, the reading of translations by hearing officers or their assistants can reinforce their negative view. It is necessary to put the stories in a right light in each of such documents. To stop reinforcement of negative view, accurate

translations and necessary annotations should be provided in each translation.

Even though language context is poorly defined in linguistics, but the accurate determination of context information in the translation environment can be made without real difficulties. The first thing one must determine is if there is any language contextual shift. For example, foreign language documents created for matters in the United States may have little to do with any foreign context. In this case, verbatim translation is the best option. If the context used by the original author is different from the context that the expected readers know, there is an assumed shift of language context.

A database can be readily built for holding context information. Language context can be classified into general context elements and context elements specific in the case. General context elements can be easily built, and they include common knowledge in legal environment, business environment, natural environment, social environment, and cultural environment. For example, general company practices under the foreign corporation law are general elements while specific mode of business used by the client is a specific contextual element. Some examples of contextual elements in the legal environment include legal system, court system, civil procedure, and hearing practice. The elements in natural environments for some countries may include provinces, major cities, and other geographic basics. Most contextual elements can be collected from public sources and loaded into a database for search. Contextual elements specific to a case can be developed only during review.

For documents which are intertwined with too many foreign contextual elements, annotations may be provided in the following priorities: (1) a context element that directly affects an issue in dispute and also directly affects the issue's resolution; (2) a context element that directly affects an issue in dispute and affects the understanding of one or more facts relating to the issue; (3) a context element that directly affects the understanding of the whole document which has some bearing on the case; (4) a context element that directly affects the understanding of some part of the document; and (5) a context element that affects immaterial words and phrases. In any specific case, it is not difficult to determine whether annotations should be provided in most text points. Also, annotations should be provided if a translation may create a false impression of some wrongdoing and invite new liabilities as a result of contextual shift.

Full text search method and "reference table" feature can be used to assist the

reviewers in identifying any of social issues. For example, if a text mentions Chinese residency as a condition for doing something, the Chinese words "residency [two words in Chinese]" should be searched. Those words are highlighted, and an explanation of residency is displayed in a reference table or an intelligent reference table (which shows those terms found in the original text). The reviewer can immediately know how those words are different from the counterparts in the target context. The reviewer then enters a note, indicating that residency is something of considerable value in the source context. This note puts the story in a right context and thus effectively prevents hearing officers from reading in American meaning by default. This method can prevent readers from reading meanings defined by the target language context in all similar cases.

I propose this extraordinary method on the basis of careful benefit and risk analysis. I can show that nothing can work better than providing foreign context information as annotations in assisting hearing officers to find truth and reach fair decisions. Every scholar can easily see a potential weakness that the annotations could be inaccurate. However, potential inaccuracies and even errors are still minor relative to the fatal errors from reading in a wrong language foundation. Thus, the proposed method is not only the only workable method but also the best method for solving this toughest linguistic problem that most scholars have acknowledged.

5. Foreign Language Context in Documents of Second Language Speakers ("SLS")

Many companies require their employees in a foreign culture to write business documents in English, and some advisers suggest that employees write English to save discovery costs. While such documents facilitate casual communications between employees and other English-speaking persons, their use in litigation can pose even bigger risks.

When the author, who grows up in a non-English language culture, writes a document in English, she essentially writes the document through translation. What the author does is to (1) form an idea to express, (2) develop native sentences, (3) develop English sentences by imitating native sentence structure, and (4) substitute native foreign words by English words by pair-wise replacements. This four-step process has a big room to introduce errors and inaccuracies. Such documents are totally comprehensible to his colleagues who know the foreign culture, verbal context, and transaction context, and would use the same process to create English documents, but are totally incomprehensible to those, including hearing officers, who read the documents from an entirely different language context. The contents of such

documents are inevitably intertwined with the foreign context. Whether the documents can be used without further treatment would depend upon who reads the documents. There are specific reasons for communication failure.

One of the specific reasons is that foreigner names, foreign places, foreign business operations in the documents can be enough headache to native English readers because they cannot get any supplemental information, cannot distinguish between facilities names, location names, object names, and human names, and do not know anything about foreign business environment and legal environment.

Another problem is that the author might use the second language or the target language to mean something that does not intend. While the author can express ideas in his native language well, the author may have some difficulties to express same in the target language. The author thus misuses words and sentence structures. One big problem is that the author may pick up certain inaccurate words consistently, and leave no vestige to reveal their misuses, those errors and omissions may become perceived facts, which cannot be corrected. Some errors may be detected by native speakers, but cannot be characterized as such due to lack of reliable evidence.

Another problem is meaning change due to cultural differences. In one example, the author may use "running dog" to mock a hired thug in a native culture, but the term is attached with a loving character in the English culture. The author may try to express a concept in the target language, but is unable to find a right way to express it, and may use a quoted native term to express it. When the author is unable to use a usage in the target language, the author may also quote a translated version. For example, a well-known English way to express an action for achieving two purposes is "one stone for two birds." A similar Chinese expression is "One arrow for two birds." If the author does not know "one stone for two birds," the author may write "one arrow for two birds" in English. While this example may provide enough similarity, other substituted quoted expressions may be incomprehensible to native readers in the target language. All inaccuracies, errors, and incomprehensible terms are written into original documents. In the worst case, they may quote foreign language terms directly in the text.

One example showing the effect of contextual shift concerns employee's sign up practice. A new employee in China needs to process employment paper work and gets a new worker identification number. Based upon original Chinese statement, the official statement in

English is "New employee will register on the first day and apply for an employee number." While this is the closest to the Chinese statement, it does not make any sense in the Western culture. Due to many differences in different cultures, this type of expression differences are common.

Word usage differences can lead to an unintended meaning. One example is "The payment to our vendor was rejected by the Internet bank due to over payment." It is hard to understand why a bank rejected a payment because of over payment. However, intended meaning is really that the bank rejected the payment because the amount exceeded the bank balance or draft limit. The author wrote this English statement by imitating Chinese language tradition. In Chinese, "付费过高" means "over payment", "over the limit" and "over the balance" and any other situations where the amount is more than a certain number. The author does not follow English tradition that insists that "over payment" and "over the limit" are for two different situations. Therefore, this statement may be changed into "the payment to our vendor was rejected by the Internet bank due to [over the limit or insufficient balance]." This kind of problems is also common.

Another class of common problems is word order, plurality, articles, and so on. The source language and the target language may use different word orders and require different information. In some foreign languages, when a concept or thing is modified by many adjectives, they may be placed in front of the word. In another language, the modifiers may be organized with different structures. Such differences can also cause all kinds of problems.

Editing this class of documents is the most difficult among all tasks in document processing. The prerequisites for editing such documents is that the translator has enough knowledge of both language contexts plus required knowledge of special technical art. In addition, the translator needs to develop certain sensitivity for identifying proper texts for annotations. The translator, who is in a super position to understand documents in any of the two language contexts, could omit annotation for texts that need annotations. The translator must be consciously aware of different effects of different contexts on the interpretation of relevant texts. Several methods may be used to make this task easy. One of the methods is a dynamic global guidelines for identifying problems. The most probably texts are the words and phrases relating to social and culture concepts, business methods, legal system, and distinctive national conditions. A database taking advantages of experts' works may be used to find texts

by searching key words and phrases. Other words and phrases, which are unique in the client case, can be collected by the entire team. It is unrealistic to expect a single translator to take care of everything.

6. Problems in Translating Troublesome Documents

Business documents are created for conducting business in the maximum efficiency at the expenses of accuracy and elegance. Documents may be created by shorthand, leave out the substance that are known to intended readers, contain grammar, usage and spelling errors, contain uncommon abbreviations, partial dates, unclear or implied times, misleading place names, and incorporate unstated external facts. A vast number of documents may be informal, casual, concise, with background and introduction omitted. Many additional factors that make the problems worse include complex background technical matters, handwriting illegibility, encoding errors of foreign language, and even omission of people, locations, and facilities names. It is often impossible to decipher this type of documents. Some examples include laboratory notes, calendar entries, informal technical reports, brief customer field reports, and database compilation.

If an original document has one or more problems, it is impossible to keep the same problems in the translation because the source language and the target language do not share the same vocabulary and same sentence structures. A source language such as Chinese does not need to indicate the number of nouns and do not require the consistency between subjects and verbs. The missing of any information, which is required by the target language, may make it even more difficult to preserve original meanings. They may differ in many concepts such as articles, noun plurality, number consistence, tenses, voice, conditional expressions, and specific usages. When a sentence is translated from a source language to a target language, necessary changes must be made based upon the substances and respective language rules. There is no way to keep the same level of uncertainties and same errors in the translation. Misspellings in English cannot be expressed in other language because they use different writing symbols; improper grammars cannot be preserved in a target language because they are governed by different grammar rules; improper use of adjectives and adverbs in the source language may be obvious, but the vestige of errors cannot be preserved in translation; casual expressions in English cannot be translated as similar casual expressions in the target language. Essentially most problems in an original document cannot be preserved with useful

information. A resulted translation may lose original meanings, contain distorted meanings, and have additional meanings, and most of the information necessary to ascertain original meanings may be distorted or lost.

If a laboratory note is about a series of experiments, verbatim translation of the note can introduce all kinds of errors. A subject mentioned in a research note may be "correctly" translated as a single person, but may be contrary to reality. If a laboratory note is about experimental design, the accuracy of the translation depends upon the proper writing-in of the plurality for various subjects. The pluralities of the subjects in the experiments are not directly revealed in any text, and nor in the forms of associated verbs. They can be inferred from experimental designs. Verbatim translation with arbitrary selection of pluralities for the subjects may result in a translation revealing an impossible experiment. Failure to make proper and necessary changes in any of many aspects can also result in confusing and incomprehensible translations.

The ability to understand substance in the source language is critical in translating troublesome documents. Source documents may contain a large number of specialty words, technical jargon, abbreviations, acronyms, informal expressions, cultural expressions, company names, people and locations, and production facilities. The translators' familiarity with those concepts is a requirement for properly translating troublesome documents. In any language, a word may have one to potentially hundreds of meanings. The first a few meanings are for common uses, but the meanings in high order numbers are generally for special trades and technical fields. The translator must identify right words. In such documents, company names are written in one to two characters, locations may be spelled as notations of one or more characters, human names may be initials, numbers, and notations, and facility names may be informal. The documents may contain words and phrases having cultural meanings and even contain local dialects. Thus, proper translation of the documents requires multiple-fields background knowledge and a full understanding of underlying technologies.

If omitted assumptions cannot be found from original documents, they may be established by reading other related documents. In translating troublesome documents, one thing that could help the translator improve translation accuracy is to review related documents. By reviewing related documents, the translator can learn relevant companies, people names, locations, facilities, and background stories, and improve translation quality. If documents are

selected and sent to a vendor for translation, the vendor is not provided with any context information, and nor is it informed of the focus of litigation. The vendor has to translate them by guess in a context-deficient setting.

When troublesome documents are verbatim translated without annotations, it is highly probable that original meanings are excluded and unintended meanings are added. Verbatim translation should not be used to translate troublesome documents if they are related to issues in dispute. The problem is worsen when language contextual shift is expected. Their meanings relevant to issues in dispute may be deciphered from studying original documents and related documents. After the documents are translated into a target language, they may become indecipherable riddles. Such translations cannot be used as critical documentary evidence. If those documents contain matters that are remote from issues in dispute, translations may be used for exclusionary purpose. If they revolve around any critical issues, they should be translated with the most possible meanings, accompanied by a construction analysis.

E. Collaborative Foreign Language Review and Translation Environment

1. Complete Translation Environment

The following figure shows a foreign language document review and translation environment. Under such a review/translation environment, every tool discussed above, and all configured and fixed database tables are accessible to all reviewers. The project manager can set up a review project, and set up configurable tables. Each of the tables contains terms in a source language, correspondent term in a target language, and definition or comment in any language in various ways. Each reviewer can search information by local, interactive and global search methods, add useful case information to the database, use personal pad's search tool to search personal notes, and use reference tables to show common, special, or relevant terms in foreign languages and/or target language. The reviewer can also see updated instructions, updated coding rules, and an index table showing updates. Every feature discussed above or in Discovery Information Management System or the integrated review-assisting system is available to each of the reviewers in translating documents.

Words and phrases which are closely related to foreign context may be entered in the term table as a type denoted by type "context." When a document contains any of those terms, the reviewer can find their definitions or comments by local search, global search, interactive search, or reference table.
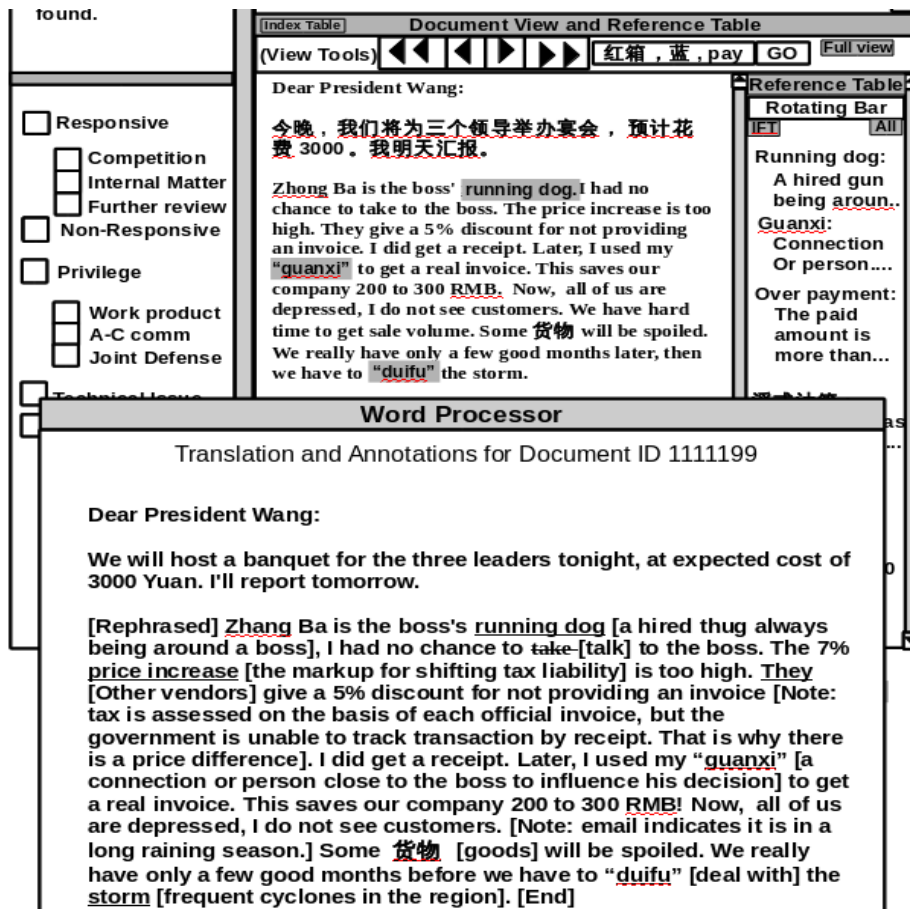
[Index Table] Document View and Reference Table

(View Tools) ◀◀ ◀ ▶ ▶▶ 红箱，蓝，pay GO Full view

Dear President Wang:

今晚，我们将为三个领导举办宴会，预计花费 3000。我明天汇报。

Zhong Ba is the boss' running dog. I had no chance to take to the boss. The price increase is too high. They give a 5% discount for not providing an invoice. I did get a receipt. Later, I used my "guanxi" to get a real invoice. This saves our company 200 to 300 RMB. Now, all of us are depressed, I do not see customers. We have hard time to get sale volume. Some 货物 will be spoiled. We really have only a few good months later, then we have to "duifu" the storm.

Reference Table
Rotating Bar
IFT   All

Running dog:
   A hired gun
   being aroun..
Guanxi:
   Connection
   Or person....

Over payment:
   The paid
   amount is
   more than...

☐ Responsive
   ☐ Competition
   ☐ Internal Matter
   ☐ Further review
☐ Non-Responsive

☐ Privilege
   ☐ Work product
   ☐ A-C comm
   ☐ Joint Defense
☐ Technical Issue

Word Processor

Translation and Annotations for Document ID 1111199

Dear President Wang:

We will host a banquet for the three leaders tonight, at expected cost of 3000 Yuan. I'll report tomorrow.

[Rephrased] Zhang Ba is the boss's running dog [a hired thug always being around a boss], I had no chance to take [talk] to the boss. The 7% price increase [the markup for shifting tax liability] is too high. They [Other vendors] give a 5% discount for not providing an invoice [Note: tax is assessed on the basis of each official invoice, but the government is unable to track transaction by receipt. That is why there is a price difference]. I did get a receipt. Later, I used my "guanxi" [a connection or person close to the boss to influence his decision] to get a real invoice. This saves our company 200 to 300 RMB! Now, all of us are depressed, I do not see customers. [Note: email indicates it is in a long raining season.] Some 货物 [goods] will be spoiled. We really have only a few good months before we have to "duifu" [deal with] the storm [frequent cyclones in the region]. [End]

FIG. 3 Server-based Translation Environment for Mixed Languages

Some example keys are "tax mitigation" "residency" "storm" "raining season" and "landlord." Each of them may carry a great deal of supplemental meanings from a foreign language context.

Tracking translations for documents becomes a simple task of retrieving data records from the translation tracking table and adding data records to the table with an option to validate data records. Translation tracking table works exactly same as any other tables except that it has additional functions for uploading and downloading both original documents and optional function for uploading and downloading translated documents.

The translation tracking database table automatically works with a data entry tool for entering new records and search tools for finding data records. When a reviewer has translated a document, the reviewer enters both basic information, file paths for the original document and the translation on a data entry form, and have the data record saved on the server and files

uploaded to the server to be stored in the database table or other designated storage space. To avoid potential duplicate work, it is preferable to create a data record for the document before starting translation so that other reviewers can know that the translator has started translating the particular document. Upon the finishing of the translation, the reviewer then uploads the finished translation into the server for immediate access by other reviewers if it is required.

When a reviewer encounters a document and needs to decide whether the document has been translated, the reviewer first checks the translation tracking table. It is preferable to open a translation tracking table or refresh an opened table to get the latest information. If the table is large, the reviewer may have to use a search function to conduct one or more searches. If the reviewer finds that a translation has been done, the reviewer can add a note in the reviewer's note, check the check box for the translation tag in the coding pane, or write a comment in any other suitable field.

The translation tracking table and the precedence rule still cannot be used to eliminate all potential problems in limited situations. One of the situations is that when a document is so large that only part of the document is relevant. For example, if an email contains more than 20 individual messages, the chain email may contain responsive information in the 10th message and the 15th message. It would be very unproductive for the reviewer to translate the entire email. The reviewer may use the address block of the 10th message as logging information. The address information in the relevant message and other characteristic words and phrases in the description field or body may be used for identification purpose.

The second class of documents requiring special handling is large databases, compilations, excel files, and large presentations, which are for the maintenance of data records. In this class of documents, any part of entries can be treated as independent from other entries. The only thing that needs to be translated is the column header and those rows containing relevant information. If there are multiple responsive rows, those rows may be placed next to each other with or without indicating omitted rows. Line numbers, page names or numbers, or approximate locations on each page should be indicated for responsive entries.

2. Minimizing The Effects of Term Multiplicity

A large number of translators and a large number of documents facilitate the creation of a large number of translations that contain confusing, inconsistent and even conflicting terms. To improve readability of translations, certain identity-related terms should be used consistently

in all translations in the case.

Term multiplicity can be eliminated for several types of terms ("target terms"). The scope of the target terms should be determined on a case-by-case basis. The target terms include a set of default target terms plus any words and phrases that are relevant to issues in dispute. The default target terms include entity names, human names, product names, material names, shipping description, facility locations, dates and times (a unique calendar system), and organizational structures etc. This class of target terms can be predetermined for all cases involving any given two languages.

The issue-related target terms can be determined by looking at the case. For example, if a critical issue in a case is a company's marketing conduct, then all terms concerning its marketing activities become target terms. A particular market activity known as fax reply may not be rephrased as fax return, fax response, returned fax, replied fax, and fax back even though they can be used in ordinary situations. If an issue in dispute is certain fruit species, then all terms revolving fruit species must be used consistently.

The major steps for minimizing term multiplicity in target terms include (1) defining the scope of the target terms, (2) creating a database table or using an existing database table, (3) adding terms to the database table by the reviewers while reviewing documents, (4) translating the documents by the steps of identifying each target term, finding its correspondent term from the database table, and using its correspondent term as the proper translation, and (5) saving the translation on the client computer, and uploading it to the server by uploading form or by email. Every function and feature discussed above can be used. Those steps can be implemented as a reference table using right database as feed data.

It is desirable that term multiplicity in other types of terms can also be narrowed down as much as possible. Considering expression diversity in any language and potential costs for achieving it, it is impracticable. The only thing that may be achievable is to set up an outer boundary for stylistic and terminological diversity. It is hoped that a group of reviewers can develop preferable translation styles under the guidance published in instructions so that they do not create completely different translations for similar documents.

3. Edit Documents Created by Second Language Speakers

There is no convention to "translate" or "edit" documents created by second language speakers in the legal industry. No one has proposed this idea, but it is a must. If such

documents are used as critical evidence or control law in the target language, the types of corrections to be corrected include (1) obvious spelling and writing errors; (2) distinctive errors in usages, expression orders, plurality etc.; (3) all quoted foreign terms such as sound equivalents, original characters or words, or other foreign-language forms; and (4) and proper adjustment to offset the effects of contextual shift.

A dedicated table or shared table for storing special terms is created holding terms such as running dog, residence, over payment.... If those terms are stored in a dedicated table, they may be designated with several types ("common," "distinctive" and "personal" etc.), depending upon nature of errors. Common usage errors in many languages by a unique second language speakers (i.e., common English usage errors by Native Japanese speakers) can be found from published works in native culture. A table can be created and their terms can be loaded into the dedicated database table or the shared table. All functions implemented in discovery information management system can be used for those terms.

In editing such documents, it is preferable to correct obvious errors first. There is no rule for defining obvious errors, but this class of errors is obvious to a native speaker. This type of errors are normally influenced by their native language. Quoted terms, which may have many forms in native languages, should be replaced by target equivalents. This is a word-for-word replacement in light of verbal context. Finally, the effects of context shift on perceived facts should be offset. This task may be the most painful and difficult one. This may require special skills, special training, systematic method, and special database tools.

The tasks of translating and editing a Chinese document is demonstrated below. The document contains an email "From: Zen Ping Zhang; To: President Wang; Sent: April 10, 2009: 01:12:30; Re: 宴会/invoice," with the following text:

今晚, 我们将为三个领导举办宴会 , 预计花费 3000。我明天汇报。

Dear President Wang: Zhong Ba is the boss' running dog, I had no chance to take to the boss. The price increase is too high. They give a 5% discount for not providing an invoice. I did get a receipt. Later, I used my "guanxi" to get a real invoice. This saves our company 200 to 300 RMB. Now, all of us are depressed, I do not see customers. We have a hard time to get sale volume. Some 货物 will be spoiled. We really have only a few good months later, then we have to "duifu" the storm.

This document is completely comprehensible in the eyes of native Chinese reviewers, but is totally incomprehensible to native English readers. A large number of documents like this

can be found in any foreign language case. In light of language context in China, the English part is edited as follows:

> Tonight, we will host a banquet for the three leaders at expected cost of 3,000 RMB. I'll report tomorrow.
>
> [Edited] Zhang Ba is the boss' running dog [a hired-thug always being around a boss], I had no chance to ~~take~~ [talk] to the boss. The 7% price increase [the markup for shifting tax liability] is too high. They [other vendors] give a 5% discount for not providing an invoice [Note: tax is assessed for each official invoice and the government is able to track each official invoice, but does not track receipt. That is why there is a price difference]. I did get a receipt. Later, I used my "guanxi" [connection or person close to the boss to influence his decision] to get a real invoice. This saves our company 200 to 300 RMB! Now, all of us are depressed, I do not see customers. [Note: the email indicates it is in a long raining season.] Some 货物 [goods] will be spoiled. We really have only a few good months later before we have to "duifu" [deal with] the storm [frequent cyclones]. [end]

This annotated text can convey its original meanings even though it is indeed a pain to read the edited document. As shown in FIG. 3, all tools discussed above are available for the translator. Real-time case information, reference table, coding rule table, foreign terms table, special terms table, and other factual tables are available to the reviewer during translation. Those tables contain the work product of language experts and work product of the entire review team from the first day of review. Each of the reviewers thus stand on the "shoulder of a superman."[2] The translated and edited document is sent to designated database or folder on a server by a file uploading form or email.

4. Proposed Method for Translating Troublesome Documents

Best translation of troublesome documents can be achieved only by ascertaining original meanings in all possible ways. Translation accuracy depends on translation approach and the translator's ability to analyze original substances. The translator needs to understand background, specific substance, and related facts, and weigh the probabilities of possible meanings. The sole judgment for translation quality is how well to preserve: (1) the scopes of terms including subjects, objects, and concepts, (2) the logic relationship between any two parts, (3) the timing of all stated and implied events, (4) the order of all things, (5) the spatial

---

2. Based upon a well known Chinese expression that a person takes advantage of a great success of a great man so that the person can speed up the process to succeed.

relationship among all things, (6) the plurality of each of the things, (7) the feelings and attitudes of the author toward to other persons and things, (8) the implied purpose of each conduct, (9) the subtle relationship between a thing and its modifier, and (10) the degrees of adjectives and adverbs. There are many additional parameters applicable to specific cases. They are referred to as "construction parameters" for translation purpose. After a document is translated from the source language to the target language, if any of those parameters are changed, the translation contains an inaccuracy or error.

Not all of the "construction parameters" listed above can be found in every document. Even perfect translations may be unable to preserve all of them. It is impossible to predict whether change of any of the parameters interferes with factual findings or application of law. The feelings and attitudes of the author toward a particular subject, person, or thing may be meaningless in many cases, but may be critical in finding willful, reckless, and hostile conduct in punitive damages claim; the plurality of items may support or repudiate a legal theory; and a different spatial and sequential relationship may effectively repudiate a patent infringement claim or patent validity position. By using the construction method, certain meanings may be ruled out based upon ascertained facts, technical or logical impossibilities, or cultural prohibition, and certain meanings can be inferred. One or more interpretation differences may affect the disposition of an issue in litigation. To reduce burden of work, the analysis should be conducted with legal issues in mind.

The nature of translating a troublesome critical document is exploring all possibilities from the native language. The translator needs to explore technical terms, specific case facts, abbreviation meanings, omission patterns, external facts, omitted assumptions, and personal writing styles, and personal distinctive error patterns. The translation environment can provide ample tools including database for technical terms, technologies basics, common usage errors for any two languages, and context database in both general terms and specific terms. The translator writes analysis report together with a translation. The translation environment with fully loaded database can dramatically reduce time and improve accuracy.

5. Tools for Processing Other Case Information.

The translation environment also provides real time review instructions and updated coding instructions. Use of illustrative coding map can effectively reduce the risk of porting different coding logic into the current project. The map should be drawn in a way to reduce

undesirable effects of importing reviewers' experience and knowledge into the current project. The information system also has many power tools. The privilege module is also built in a real time collaborative model which allows the reviewer team to share work product in real time. Other tools include identity data search algorithm and name lists processing web modules, which take editing tasks away from the reviewers so that they can use the time on more important things.

F. Strategy to Contain Translation Costs

There are many reasons for high document review costs. When documents have foreign languages, translation time is responsible for a good part of discovery costs. There are several causes for high translation costs. One of the causes is a large number of duplicate documents that cannot be eliminated. For example, a particulate document is created by distributing a draft to a group of employees for review and the drafter gets each of edited documents for inclusion in the final version. This process can be repeated many times in several months up to several years. This practice results in hundreds to thousands of copies that are not identical. All those copies may be assigned to a large number of reviewers. While each reviewer can identify the most inclusive copy within the reviewer's document range or folder, there is no easy way for the whole reviewer team to figure out which one should be translated. Tens of reviewers might tag identical documents for translation. The project manager or litigation attorney, especially if he cannot read the source language, cannot determine whether those documents are identical based upon reviewers' notes. If some of the documents have been translated, the attorney cannot tell whether the translated documents are identical because all translations look somehow different due to "term multiplicity" and "structure multiplicity." When there are several litigation attorneys and a group of reviewers, they can routinely spend a great deal of time to figure out which documents have been translated and which need to be translated. This translation task tracking method is the best solution.

Good translations require a great deal of time and even an unlimited amount of time. Translation time in terms of dollar amount per word can vary by many folds, depending upon nature of original documents, accuracy requirements, translator's knowledge, and translation philosophy. A decent translation of a document requires more knowledge than what is required to paraphrase statements in the document and far more than what is required to code the document. If accuracy is not required and the subject in the document is very general, Google

and Ping translation tools may be good enough and the reviewer just takes a look at the translation to ensure that the resulted translation does not contain harmful errors and inaccuracies. If accuracy is not required and the reviewer has good background, the translator can translate documents in nearly typing speed, and still achieve moderate accuracy. If the documents are highly technical and the reviewer lacks required relevant knowledge, the reviewer has to struggle, but still cannot deliver required accuracy. If the documents contain highly technical matters and also require the highest accuracy, the documents may demand the highest skill levels and require a great deal of time. If the documents are highly technical, concerned with critical issues, and contain some problems such as bad handwriting, casual notes, missing pages, and errors and omissions, it would be a research project. A version of translation may require several rounds of amendments. A challenge by the opposing party may require further amendment. When the document pool contains a large number of hot, technical, and troublesome documents, and there is no workable method for managing work flow, translation costs can be prohibitory. The translation environment equipped with a plurality of database tables and intelligent reference-generating feature can dramatically reduce the time for conducting endless searches and endless guesses.

One of factors affecting translation costs is translation method. Translation methods include verbatim translation and translation by meaning. Each method can include various subtle differences. Verbatim translation can be performed much faster because the translator does not need to fully understand original content. It can be done by sentence-by-sentence replacements according to different grammar rules. The risk is that the resulted translation may significantly depart from original meanings. Translation by meaning is more difficult. If the translator does not have the knowledge of underlying art, the translator has to learn it during translation. If the subject matter is very complicated, this learning process can take a great deal of additional time, and yet the translator still can make mistakes that a well trained translator in the art can immediately know. There are also situations where subject matters are so complex that they are not for laypersons to translate. If the translator knows underlying art, the translator can learn specific subject more quickly, and can have a better chance to avoid making mistakes. The trained translator is in a much better position to ascertain implied assumptions, incorporated external facts, original mistakes and omissions, and uncommon expressions. The translation environment is particularly helpful to translators who try to preserve the original meanings.

Another factor that can also contribute to high costs is the total lack of tools for translations. On some review sites, it is unrealistically expected that translators know everything in two languages. When a foreign language is so big, no one can know everything. Chinese is a very big language that has evolved for more than four thousands years. It has such a long evolving history that it requires "translation" for readers in different eras. It has a large numbers of characters and a huge number of ways of expressions that many writings have "translations" for readers in different technical fields. On top of that are rich cultural, social, political, historical, and geographical elements. Even just human's names, location names, and company names can totally disrupt work flow to a halt. No one knows all official counterparts of thousands to millions of company names, location names, and human names. The translator may run into something that requires mini research. If separate public Internet access is not provided, the translator has to use sound translations rather than officially recognized counterparts, and may have to spend more time to fix inconsistencies and errors after the translator knows better translations. If no tool is provided and the translator cannot understand words and phrases, the translator has to consider whether the problem is an implied assumption, incorporated external fact, an original mistake, or unusual expression. If no help is available, the final try is to make a best guess under the circumstance.

Foreign language input method affects translation efficiency in some cases. Foreign language input may be necessary for searching foreign words, retrieving foreign documents, conducting backward translation for verification, and creating foreign text in translations. The input method in Windows Operating System has limited functions. Input methods for character-based languages are completely different from the input method for any letter-based language. A reviewer may be good at using one particular input method, but cannot use other input methods. Their productivity in inputting foreign language texts may differ greatly. The availability of language input methods for those languages can affect work efficiency in certain cases. If backward translation is necessary, a reviewer may be able to perform very well by using one input method, but cannot work at all by using another input method.

The number of foreign languages in documents also affects review productivity. Some documents may be created in more than two languages in alternate lines or sections. Such documents need many rounds of reviews. Strictly speaking, competent review cannot be performed by multiple independent reviews because such review is unable to ascertain combination effects. Each of the sections may be non-relevant when the sections are read

separately, but their combination may present significant issues. Demanding one reviewer to review multiple languages is also questionable because there is no guarantee that a single reviewer is really competent to review several languages in native levels. If a document in many languages is a product of one single author, it is also questionable whether the author is able to convey intended meanings in the languages. The translation environment supports as many foreign languages as practicable.

The amount of case information always affects translation costs because each of the reviewers must learn it. When the amount of information is doubled, the time spent on the learning process for all reviewers is also doubled. The large amount of case information, numerous file types, and common technical problems may be intertwined to further increase translation costs. Poor review plans, failure to provide background knowledge, incomplete and confusing review instructions, and missing support applications on review computers are among other factors that contribute to high costs.

Great efforts have been made to reduce discovery costs in the review industry. Certain search and file elimination methods may disrupt cross-document verbal context, and make some critical documents unavailable. Some search methods used for processing documents make translation tasks more difficult or force translators to make best guess. Some computer search methods can reduce document review pool by as much as 80%. This may reduce the number of documents to be translated. The deduced size of the document pool can reduce the total production cost, but it may affect accuracy of translations if it affects verbal and transaction context. The method may make some translations incomprehensible.

G. Translation Performance and Translators Qualifications

To improve document translation accuracy, it is necessary to determine a suitable measure for translation accuracy.

Translation performance depends on three things: the selection of documents for translation, the ability to preserve original meanings, and overall translation costs. Selection of documents depends on nature of case, issues in dispute, litigation strategy, and document composition. The document selection process is affected by a battery of practical difficulties arising from division of work and division of documents. The quality of translations is ultimately judged from their performance in achieving the client's final goal. If translations give the client the best chance to win its case, the quality is excellent. If they give the client a fair chance to

win or defend its case, the quality is fair. If they ruin the client's chance to win its case, the quality is bad.

Within the bound of adhering to facts, anything may be translated in many different ways. By using words and expressions, the translator may exaggerate the stories, accurately reflect the stories. The translation may express the stories in a best light or worst light. This is similar to what attorneys present facts in court. A good translation, which is produced by the client, should not put the stories in a worse or worst light. Performance measure should take into account how well translations address matters important to legal issues. Many of those performance parameters cannot be directly measured on review sites.

 Unlike document review where coding performance is limited by the reviewer's source language skills only, the accuracy of translations is limited by the ability of the translator in any of the two languages in the relevant field. It can be safely assumed that no translator is competent to translate critical documents that are created by particular authors in particular contexts. Due to prevalent knowledge disparity among individual persons in any culture, it is highly improbable that a translator somehow has language skills, technical skills, special knowledge, and context knowledge that can match those of a particular author who creates documents. The only remedy to cure translators' knowledge deficiency is a systematic method, use of mass database tools, and a collaborative environment.

In translating highly complex documents, the length of translation experience may have little relevancy. If a reviewer lacks relevant technical knowledge and technical ability necessary for understanding particular types of subject matters, the translator translates the documents by word-for-word replacements. Such translations are confusing and are incomprehensible in light of both technical context and the target context.

The relative strengths of translators in two languages affect translation quality in different ways. If the translators' weakness is in the source language, the translator may poorly understand original substance but still can express misunderstood substances in elegant target language. Misunderstanding finds its way to the final translation. If the translators' weakness is in the target language, the translator may be able to understand the original document correctly, but is unable to express what the translator understands in the target language. The translation thus has bad English on its face. Both contain distortions and inaccuracies.

In the review industry, there is a prevalent misunderstanding of translation quality. Some lawyers pay too much attention to the looks of translation work product. Original documents are created in a different culture for efficient communication, the best and most accurate translations can be anything but elegant writings in the target language. The awkwardness of translations is in part due to the constraints imposed by the original expressions and the foreign context. Elegant brief-style translations inevitably contain distortions and inaccuracies. Any policy to solicit brief-like translations forces translators to rewrite stories, which may have unpredictable effects on outcome of the cases.

High accuracy is achieved if a translator is able to understand original documents accurately and express his or her understood matters accurately in the target language. One method for achieving the best accuracy is that a document is translated jointly by two translators, one with native fluency in the source language and one with native fluency in the target language. The first translator translates the document into a draft translation as accurately as possible. The second translator then edits the draft, and, if necessary, rewrites the translation. In this process, they should discuss how to improve troublesome text in the best way. This method may be used to translate troublesome and critical documents.

Translation speed in terms of words per hour may vary by several to hundreds times. As in document review cases, translation quota will not achieve any good end unless mistakes, inaccuracies, risk exposure, and self-incriminatory misstatements do not hurt the client. There is a fundamental difference between widget manufacture and litigation production. In manufacturing, the total revenue is always based upon the quantity of produced wedges. In litigation, the total production costs depend upon work volume, but the outcome of the case does not. In many noted leading cases, their outcomes hinge on a proper handling of a small set of key documents in support of or refuting disputed facts. The client may save litigation costs at the expense of a total defeat and exposure of additional or future liabilities. This distinctive nature requires that document reviewers try their best to increase the chance to win. Any attempt to discount efforts for the sake of meeting daily quota and achieving performance ratings may jeopardize the prospect for victory. Reviewers' passion for protecting client's interest is an absolute requirement for getting the best chance to win.

H. Use Translation Algorithms and the Translation Environment Combination

Computer generated translations may be directly used when mistakes, inaccuracies, risk

exposure, and even self-incriminatory misstatements do not hurt the client. Computer translation algorithms may be used to create documents' main ideas for the purpose of responsiveness determination in limited cases, and used to generate translation templates for edits. By using such templates, the translator can save time. If computer translation algorithms are used to identify central idea, such translations cannot be produced as party evidence because they may contain improper substance.

There are many reasons why computer algorithms cannot do a good job. Current algorithms are unable to (1) understand background technologies and technical matters, (2) associate various parts of language context to make an intelligent judgment, (3) associate different parts of materials in the same document, (4) properly treat assumptions, obvious expression errors, obvious omissions, and improper abbreviations, (5) appreciate informal expressions and all kinds of secondary meanings such as connotation, implication, sarcasm, and hidden messages, (6) detect inherent properties of matters, things, and events, (7) understand human emotion, malice, and intention, (8) make connection between two or more things by times, persons, event, or concept, and (9) recognize handwritten notes and contents on most drawings.

The above list is not intended as a complete list. In some situations, even skillful reviewers may disagree on how to interpret contents in troublesome documents, computer translations may be used only in cases where articles are written in a plain style without any of those discussed substances. One difficulty is that the documents do not contain any indicators whether they have any of those issues and encoded secondary messages. Secondary messages may be in documents concerning organizational power struggle, factional disputes, and matters relating to politics, and this is especially common in certain Eastern cultures. A secondary message can be encoded into a document by citing a historical character, famous event, well-known story, or special date.

If document security and document confidentiality are not an issue, one good method to improve efficiency is using a three-step process: (1) translating a document by internal or public computer algorithm such as Google Translate, (2) searching original texts for all classes of the keys from the database tables, and (3) optionally, searching the terms in the translations against a correspondent database. The words and phrases reflect the work product of many experts in the relevant languages, and the cumulative work product of the entire review team.

The highlighted terms and their explanations help the translator know most suspected problems immediately. This process can relieve the translator from spending time for typing text in the target language and constructing simple sentences, and giving the translator more time to focus on important matters. This method can dramatically reduce the number of searches and guesses while maintaining the same level of accuracy, or help the translator achieve the highest accuracy. At the optional step, the translator can conduct a search in the translation against a correspondent database to have all suspected terms highlighted. This feature can be used to verify whether English words and phrases are used inconsistent with original native counterparts.

Conclusion

The server-generated environment can relieve translators from conducting endless searches or endless "guesses," which are common in translating corporate business documents in large discovery projects. The system loaded with database and shared case information can help the whole review team dramatically improve overall review and translation efficiency. The protocol can help translators produce translations that are comprehensible to those who do not know the source language context. The system, which can run as an independent super translation tool, is stable, secure, and pose no risk to any other parallel review systems. It can meet the search needs for a vast majority of searches in a closed secure network, private cloud, and public cloud.

Product Information

The server application software is available for non-exclusive limited licensing. The system can be deployed on any LAN server, regional server, and private and public cloud. The owner reserves all rights in the software, system data, system forms, data templates, source code, and forms and layouts, whether they are patented or not. For product information, please contact john at jwu9999@gmail.com or 202-560-3000.