The LANGUAGE Center

190 Rt. 18, Suite 203
East Brunswick, NJ 08816
phone: 732-613-4554
fax: 732-238-7659
tlc@thelanguagectr.com
www.thelanguagectr.com

The One Source for a World of Translation Services

Glossary of Industry Terms

Accent Mark. A mark or sign placed under, over or through a character which indicates a modification in its phonetic value. See also Diacritic.

Alignment. The process of ensuring that each segment of a source text correctly corresponds to its target segment.

ANSI. American National Standards Institute is an organization of American industry groups that work with other nations to develop standards.

ASCII. The American Standard Code for Information Exchange is the worldwide standard for the encoding used by computers to represent all the lowercase and uppercase Latin letters, numbers, punctuation marks and other symbols.

Back Translation. A back translation helps to determine if the original meaning has been preserved in translation and is the process of translating a document back to its original language in a literal manner.

Bidirectional Text (bidi). A mixture of characters where some are read from left-to-right and others from right-to-left. Bidi refers to an application which allows for this variance.

Big5. The name of the Chinese character set and encoding used extensively in Taiwan.

BMP. A standard bit-mapped graphics format used in Windows.

BinHex. Binary Hexidecimal is a widely-used encoding scheme that coverts binary data into ASCII characters. Especially common on Macintosh platforms.

CAT. Computer-aided translation (CAT) describes computer technology applications which assist with translating text from one language to another.

Central and Eastern Europe (CEE). While not precisely defined, this term usually refers to the European countries east of Germany and south to the Balkans.

Character. The smallest component of written language that has semantic value. A printed or written letter or symbol. In computing, binary code used to represent a letter or symbol.

Character Set or Charset. The mapping of characters from a writing system into a set of binary codes such as ANSI or Unicode.

CMYK. Cyan Magenta Yellow Black. A color model in which all colors are described as a mixture of these four process colors.

Code Page. A table which defines the numeric index (computer code point value) associated with each character from a specific set. Each character in a code page has a numerical index.

Colloquialism. An utterance belonging to ordinary, everyday speech characteristic of informal, spoken language. Colloquialisms are often short-lived, fad sayings.

Consecutive Interpretation. A form of interpretation where a sentence or phrase is spoken, the orator pauses, then the interpreter repeats in the target language.

Content Management System (CMS). A system used to store and subsequently find and retrieve large amounts of data. Content management systems were not originally designed to synchronize translation and localization of content, so many work in conjunction with globalization management systems (GMS).

Context. The environment or setting in which an utterance occurs. There are various contexts which are crucial for the translator to be aware of. The immediate linguistic context consists of the words, phrases and sentences surrounding the utterance in question.

Contraction. The approximate amount a target language will decrease in volume over that of the source it is being translated from, typically expressed as a percentage.

Controlled Vocabulary. The standardization of words which may be used to search an index, abstract or information database. There is usually a published listing or thesaurus of preferred terms identifying the system's vocabulary.

Copywriting. This process involves taking creative liberties that do not exist when translating a legal document or user manual and is typically a function of time as opposed to volume of words. Ad copy and marketing verbiage are typically copywritten.

Desktop Publishing. The process of using a personal computer to produce complex layouts and high-quality printed documents.

Diacritic. A mark or sign placed under, over or through a character which indicates a modification in its phonetic value. See also Accent Mark.

Dialect. A variety of a language used by people from a particular geographic area.

Diphthong. A complex speech sound or glide that begins with one vowel sound and gradually changes to another within the same syllable, such as coin, loud and side.

Disambiguation. The process of rewriting or reconstructing a sentence so that one of several possible meanings is singled out.

Dynamic Content. Information on web pages which changes automatically based on database content or user information. Most URLs ending with .php, .asp, .cfm, .cgi or .shtml are serving dynamic content. For example, results generated by a search engine in response to a query are dynamic content pages.

DLL. Dynamic Link Library. A file that contains executable functions or data for applications.

Double-Byte. Refers to character sets that require more than one byte of information to express each character. Typical for languages such as Chinese and Japanese.

DPI. Dots Per Inch. A common measurement of resolution.

e-learning. The use of internet technology for learning outside of a physical classroom.

EMEA. Short for Europe, Middle East and Africa, a term used in commerce to signify markets or activities in those regions of the world.

Encoding Scheme. Rules for assigning numeric value (computer code point values) to characters. Encoding is a method by which a character set is turned into computerized form for transmission and preservation.

Escort Interpretation. A casual form of consecutive interpretation. Tour guides or those showing people around usually act as or utilize escort interpreters.

Expansion. The approximate amount a target language will increase in volume over that of the source it is being translated from, typically expressed as a percentage.

FTP. File transfer protocol (FTP) is a common way to move files between computers.

Full Match. A source text segment which corresponds exactly (100%) with a previously translated segment in a translation memory (TM) tool.

Fuzzy Match. Refers to instances when a sentence or phrase in a translation memory (TM) is similar (but not a 100% match) to the sentence or segment being worked on. A TM tool will typically calculate the degree of similarity as a percentage.

GIF. Graphics Interchange Format. A popular bitmapped graphics file format that supports 8-bit color (256 colors) and is widely used on the Web, because the files compress well.

Gist Translation. A less-than-perfect translation typically performed by either machine translation or a process whereby limited to no human proofreading and/or editing is performed subsequent to initial translation.

Globalization Management System (GMS). Focuses on managing the translation and localization cycles and synchronizing those with source content management. Provides the capability of centralizing linguistic assets in the form of translation databases, leveraging glossaries and branding standards across global content.

Glossary. In a translation and localization context, a glossary is a list of source language terms that are paired with a list of corresponding terms in the target language.

Glyph. The shape representation or pictograph of a character.

Hanja. Chinese characters (Hanzi) that are used in Korean are referred to as Hanja. It refers to those characters borrowed by Koreans and incorporated into the Korean language with a changed pronunciation. Unlike the Japanese Kanji, which has altered and simplified many characters, Hanja are almost entirely identical to modern traditional Chinese Hanzi.

Hanzi. Refers to the Chinese characters (or Han characters) used in the written forms of the Chinese language, and to varying degrees in the Japanese and Korean languages.

Hiragana. A flowing phonetic subscript of the native Japanese writing system. In Hiragana, all of the sounds of the Japanese language are represented by 50 syllables.

Homograph. One of two or more words that have the same spelling but differ in origin, meaning and sometimes pronunciation. An example is *wind* (weather) and *wind* (activity).

Homophone. A word which has the same pronunciation as another but a different meaning, derivation or spelling, such as *there* and *their*; *foe* and *faux*; *time* and *thyme*.

HTML. Hypertext Markup Language (HTML) is a markup language that uses tags to structure text into headings, paragraphs, lists and links, and tells a web browser how to display the text and images on a web page.

Ideographic Language. A written language in which each character represents an idea, concept or other component of meaning, rather than pronunciation alone. Japanese Kanji, Chinese Hanzi and Korean Hanja are examples of ideographic writing systems.

ISO. The International Organization for Standardization (ISO) is a network of national standards institutes from 145 countries working in partnership with international organizations, governments, industry, business and consumer representatives.

IVR. Interactive Voice Response is telephone technology whereby someone using a touch-tone telephone interacts with a database to either acquire information from it or enter data into it.

Java. A general purpose programming language with features that make it well-suited for the web. Java code can be run on most computers.

JavaScript. An open-source scripting language for the design of interactive websites, it can interact with HTML source code, enabling web developers to use dynamic content.

JPEG (JPG). Joint Photographic Experts Group is the committee which sets file format standards for graphics. The JPEG file format is compressed, with some loss of quality during compression and a popular web format given the generally small file size.

JSP. Java Server Pages (JSP) have dynamic scripting capability that works in tandem with HTML code, separating the page logic from the static elements — the actual design and display of the page — to help make the HTML more functional.

Kanji. The name for Japanese ideographic characters that were derived from Chinese origin. Despite the existence of over 13,000 kanji characters, these alone are not sufficient to write Japanese. Hiragana characters are also required to express grammatical inflections.

Katakana. A phonetic script that renders the sounds of the Japanese language into 50 syllables. Katakana and Hiragana both render the same syllables, but Katakana is comprised of squares or angular lines and is used largely to spell words borrowed from other languages, while Hiragana is cursive and used more frequently to spell Japanese words.

Kerning. The adjustment of space between pairs of letters to make them more visually appealing, normally applied to individual letter pairs in headlines or large type.

Leading. The space between lines of type, generally measured from baseline to baseline and expressed in points. The name leading is derived from the days of hot metal type when strips of lead were placed between lines of type to produce spacing.

Leveraging. The amount of previously-translated text that can be reused or recycled.

Lexicon. A lexicon is the collection of words and word parts used by people speaking a particular language.

Ligature. Refers to a glyph that is created when two or more characters are combined to form a single, new typographical character.

Lingua Franca. A language that is adopted as a common one between speakers of different native languages.

LMS. A learning management system (LMS) is software that automates the

administration of training events.

Loanword. A word or phrase adopted from another language with little or no modification.

Machine-Aided Translation (MAT). Computer applications which assist in the translation of text from one language to another, based on the concept of translation memory (TM) and the reuse of previously translated terms and sentences.

Machine Translation (MT). Technology that translates text from one human language to another, using terminology glossaries and advanced grammatical, syntactic and semantic analysis techniques.

Metaphor. A figure of speech where a word or phrase that ordinarily designates one thing is used to designate another, therefore creating a comparison between two seemingly unrelated subjects. Metaphors often cannot be translated literally.

Morphology. The branch of grammar which studies the structure or forms of words.

Multi-Language Vendor (MLV). A language service provider that handles more than one language or language combination.

Multilingual Workflow System (MWS). A computer program which creates an environment that supports and orchestrates a range of activities that facilitate the development of multilingual products. An MWS should contain a globalization management system (GMS) for managing multilingual content, along with translation memory (TM) and machine translation (MT).

National Language Support (NILS). A function that allows an application to set the locale for the user, identify the user's primary working language and retrieve information representing times, dates and other information which is formatted correctly for the specified language and locale, and includes support for keyboard layouts and language-specific fonts.

Natural Language Processing (NLP). A main focus of computational linguistics, the aim of NLP is to devise techniques to automatically analyze large quantities of transcribed or written text in ways that parallel what happens when humans perform this task.

NET. Microsoft platform for applications that work over the internet.

OCR. Optical character recognition is the translation of optically scanned bitmaps of printed or written text characters into character codes such as ASCII. Most OCR systems use a combination of hardware and software to recognize characters.

OpenType Fonts. OpenType fonts are cross-platform, self-contained files with advanced typographic features such as glyph substitution.

Orthography. A language's writing system, essentially the same as an alphabet.

PCX. Paintbrush Exchange. A bitmapped graphics file format that handles monochrome, 2-bit, 4-bit, 8-bit and 24-bit color.

PDF. A file format created by Adobe Acrobat, primarily for read-only use with the free version of Adobe Reader.

Phonology. The part of linguistics that deals with systems of sounds especially in a particular language.

Pixel. Derived from the words "Picture" and "Element", a pixel is the smallest element of a display which can be assigned a color.

Plug-ins. Modules that add a specific feature or functionality to a larger system or application.

Pretranslation. Involves the preparation of files for translation whereby existing files already contain related segments of previously translated data. Only 100% matches are replaced, with the result being a set of files ready for translation that contains both source and target language terminology.

Project Management (PM). Project management (PM) is the systematic planning, organization and controlling of allocated resources in order to meet project cost, time and performance objectives. PM is typically a process of performing focused, non-repetitive, time-sensitive, project-specific activities.

Pseudo-translation. Basically a "test run" in which you seek to replicate the translation process rather than actually produce a translation. A text string is subjected to a translation-like process produces a new string. The text string is frequently changed as a result of this process, so pseudo-translation is done to illustrate potential problems that may occur when the translation is actually done.

Quality Assurance (QA). All planned and systematic actions necessary to provide adequate confidence that a product or service will satisfy given requirements for quality.

Radical. The root or base form of a word. The building blocks of Chinese characters, of which the most common set contains 214 radicals, themselves, composed of strokes.

Rich Text Format (RTF). A file format standardized by Microsoft for creating formatted text files. Unlike basic text files, RTF files include information such as text style, size and color.

Segment. Units that are a result of the parsing process translation memory tools perform. Segments can be words, phrases or sentences and allow for useful translation correspondences between source and target languages.

Semantic. Part of the structure of language, along with phonology, morphology, syntax and pragmatics, which involves understanding the meaning of words, sentences and

texts.

Simile. A figure of speech whereby one subject is compared to another, and is frequently marked by use of the words like or as, for example, "The snow was like a blanket."

Single-Language vendor (SLV). A language service provider that only works with a single language or language combination.

SGML. Standard generalized markup language (SGML) is an international standard for information exchange that prescribes a standard format for using descriptive markup within a document, defining three document layers: structure, content and style.

Simplified Chinese. A Chinese character set used in mainland China and Singapore, modified to be written with fewer strokes per character.

Simultaneous Interpretation. Interpretation that occurs at the same time the source language is being spoken. Typically requires equipment in addition to interpreters.

Source Language. The language from which translation or adaptation occurs.

Streaming. Streaming allows a computer user to see and hear an audio/video file as it is being transferred. Video is usually sent from prerecorded files, but can be broadcast live.

Supply Chain Management (SCM). An electronic alternative to the traditional paper chain, enabling participating suppliers to access up-to-date company information and enabling companies to better manage and track supply and demand.

Target Language. The language to which translation or adaptation occurs.

Terminology Manager. An application tool that assists in the translation of text from one spoken language to another.

TIFF. Tagged Image File Format. A widely used bitmapped graphics file format that handles monochrome, gray scale, 8-and 24-bit color.

TMX. Translation Memory Xchange (TMX) is an open standard, based on XML, which has been designed to simplify and automate the process of converting translation memories (TMs) from one format to another.

Traditional Chinese. A Chinese character set used everywhere except mainland China and Singapore. This set is consistent with the original Chinese ideographic form that is several thousand years old.

Translation Memory (TM). A special database that stores previously-translated words, phrases and sentences which can then be reused. The database matches

source to target language pairs.

Transliteration. To write or print a letter or word using the closest corresponding letters of a different alphabet or language. A systematic way to convert characters in one alphabet or phonetic sounds into another alphabet.

Unicode. The Unicode Worldwide Character Standard is a character encoding standard used to represent text for computer processing. Originally designed to support 65,000 characters, it now has encoding forms to support more than 1,000,000.

URL. Uniform Resource Locator or URI (Uniform Resource Identifier) is a short string that identifies resources on the web such as documents, images, files and services.

Usability. The ease users experience in navigating an interface and locating information.

UTF-8. Unicode Transfer Format is encoding form of Unicode that supports ASCII and covers the characters of most languages in the world.

Vector-based. Refers to software and hardware that use geometrical formulas to represent images (same as object-oriented graphics).

XLIFF. XML Localization Interchange File Format (XLIFF) is an open standard that was specifically designed to support the localization of data and has features for updating strings, revision control, marking different phases of the localization process, word count calculations and provision of alternative or suggested translations, among others.

XML. eXtensible Markup Language, a programming language/specification, is a pared-down version of SGML, an international standard for the publication and delivery of electronic information, designed especially for web documents.