Hypothesis on Root Language

Human beings' research on their language seems to be the longest and most varied. At the same time, it appears to be a less successful discipline. Although various schools of thought emerge endlessly, the research angles and objects are also diverse, and there are some thought-provoking research results from time to time, there does not seem to be an overwhelming theory that can lead the study of language so far. On the contrary, the development of mobile applications in the field of language is thriving, and refreshing systems and functions are constantly coming out. As a result, vulnerable basic theories are difficult to support powerful applications, which leads to the problem that application performance is still unsatisfactory no matter how the hardware system develops.

In terms of language applications, the most eye-catching and least satisfactory is the translation system. It is precisely because language research cannot provide sufficient support that translation systems can only be constructed through unreliable grammatical analysis or inefficient repetitive training methods. When within the same language family, this method can still achieve a certain quality of translation, but in the case of translation cross the language families, its quality can only cope with simple applications, and there is still a large distance from professional manual translation. The main reason is that the exploration of language by these systems only stays at the surface of the text, but cannot go deep into the level of thinking behind the language. In fact, in the schools of language research, many of them involve the relationship between language and thinking, trying to dig out something essential in the language through in-depth exploration of this relationship, but this kind of research involves some relatively blurred areas of psychology, neurology, and so on. So it is difficult to provide definite conclusions, thus unable to provide theoretical support for fundamentally improving the quality of the translation system.

There have always been many different theories about the relationship between thinking and language, but it is undeniable that thinking and language are closely related. Based on this, I propose a hypothesis here. If it is true, it will allow language research to eliminate the interference of psychology or other disciplines, bypass many unsettled underlying problems, and thus provide brand new theoretical support for a translation system directly at the language level. This would make a professional-class machine translation system possible. I call this the hypothesis on root language.

The intention of root language:

One of the aims of proposing the root language is to hopefully find out the common things between languages by conducting an in-depth exploration of language beneath the syntactic level; on the other hand, it's supposed to be focused on language, that is, to avoid the interference of psychology and other disciplines. For this reason, it is necessary to put aside the discussion on the thinking level while exploring the language at the thinking level, so that the underlying structure of the language, that is, the internal presentation form of thinking, namely the language/logic mapping of thinking, should be taken as the object of observation and exploration. Thus, the root language is defined.

The possibility and inevitability of the existence of the root language:

Human minds communicate with each other through language, and different languages do not mean different minds. After all, no matter how different the languages may be, the exchange of ideas between

human beings is always possible in the end. Moreover, anatomy has failed to detect any significant differences in the "hardware" used for thinking between different populations. This means that the human mind is essentially the same. Any expression in one language has an exact corresponding expression in the same context in any other language. Therefore, if there is an inner language dedicated to thought, then it can be assumed that this language is unified. In the exploration of the relationship between thinking and language, many theories believe that human beings use language to think. Some of them believe that the language used by thinking and the language of external expression are the same. Its representative is J. B. Watson, the founder of behaviorism. He believes that thought is language, and language is thinking with sound ("I think we ought to make the term 'thinking' cover generally all implicit language activity and other activity substitutable for language activity." IS THINKING MERELY THE ACTION OF LANGUAGE MECHANISMS? John B. Watson (1920)). Others argue that thinking takes place in a unique language (Jerry Fodor argued in "The Language of Thought" (1975) abductively: our current best scientific theories of psychological activity postulate Mentalese; we, therefore, have good reason to accept that Mentalese exists.). If a child of an ethnic group grows up in another one, his language and behavior are no different from those of his ethnic group, even though he has the thinking "hardware" belonging to another ethnic group. It is conceivable that this kind of natural intercommunication is only possible if there are no major differences in human thinking. This also indirectly proves that the existence of the root language is inevitable.

The root language is characterized by:

1. This is a common basic language based on the common physiological structure of human beings.

2. This language itself has latent traits and operates only on the level of thought.

3. The root language is projected into different real languages through the language conversion mechanism, the so-called language lens.

Root Language Structure Exploration:

Root language contains

1. Various objects and their variants acquired by the senses and their derivatives;

2. The behaviors of those objects;

3. Various delimitations (the above three are tangible contents);

4. The implication and scene assignment of the expression block (such as questions, sarcasm, humor, poetry, and adaptability to different occasions, etc.)---this part is intangible content.

Features of the root language:

It can fully contain all language information, including literal information and all kinds of implicative information beyond literal meaning.

The root language should have the most concise structure:

1. On the basic layer are the objects and behaviors and the direction of these behaviors;

2. The definition layer is various definitions for the basic layer;

3. The implication and scene layer contains various non-literal attributes and parameters.

But compared with a language in the strict sense, it does not have the extra rules of grammar but is only established through logical relations. Naming it as a root language is to include it in the language category, rather than being the research object of other disciplines.

Exploration of the presentation of the root language:

In fact, the root language is a derivative one, using symbols or numerical codes to represent the substantial contents of the language and their relationship, as well as the savor contained in the language, being able to fully reflect all the explicit and implicit contents of the natural language, while the artificial rules in the natural language are stripped off.

All contents in natural language that can be clearly delimited (i.e., objects and actions) are defined as real elements of the root language, represented by symbols, numerical codes, or Latin words. All the rest (i.e., direction, relationship, degree, implication, scenes, etc.) are defined as imaginary elements, expressed by numerical values or symbols.

The root language should be as simple and clear as possible so that the holographic export can be realized by taking full advantage of existing technologies such as text mining and information extraction.

Difference between root language and Mentalese:

The difference between root language and Mentalese is that root language does not involve thinking itself at all, and is therefore a "pure" language. Mentalese is more understood as a mixture of thinking and language, so discussions on it often have to go deep into the thinking level. This distinction is important because it ensures that the study of the root language is limited to the language level.

Conversion between root language and output language:

The root language contains all the information that people want to express. However, when the root language is converted into a natural language, part of the contained information will appear in an implicit (illocutionary) manner. As for which information is presented in an implicit manner, it depends on the information-carrying capacity of the language, the individual's language control abilities, and the need for implicit presentation in specific situations.

As mentioned above, human thinking is communicable. Therefore, when one natural language is converted into another natural language through the root language, all the information contained in the language can be completely transmitted. However, Due to the different information-carrying capacity and methods of different languages, during the conversion process, some content expressed in an explicit (locutionary) way in one language may need to be expressed in another language Implicitly stated, otherwise it may need to be explained with additional text.

The ultimate requirement for language conversion (translation) is the equivalent expression under the same situation or congruent conversion, that is, the target language translation should fully convey all the explicit and implicit information conveyed by the source language in the same way for the same context (explicit-explicit, implicit-implicit). In other words, the translation in the target language should be able to produce an exactly equal transfer effect for the same context. Film dubbing is a good example of this congruent transformation. In conversion practice, due to the different uses of the converted

(translated) text, the requirements for the congruence of conversion are also different. For example, in the conversion of technical text, almost all information needs to be transmitted in an explicit way to eliminate any possible ambiguity and uncertainty, thereby avoiding error, distortion, and loss in information transmission to the greatest extent. In other cases, such as poetry, it is necessary to fully convey certain information beyond the literal, such as poetic imagery, meter, rhyme, or even elegance grade.

Language Lens Assumptions:

Based on the relationship between root language and natural language, that is, the correspondence and mutual convertibility of the information contained in them, it is conceivable that there is an internal conversion mechanism between root language and natural language, or language filter. If the filter can be "replicated", the root language containing all the information to be expressed can be faithfully and undistortedly projected into a real language manually or automatically.

Moreover, a specific language can also be projected in the opposite direction to the root language through the same language conversion mechanism, and then, faithfully and undistortedly projected into another real language through another language-specific conversion mechanism.

In other words, the root language is obtained by stripping off the various artificially formulated rules of the real language for the realization of interpersonal communication and filtering out the explicit and implicit ideographic content without omission.

The feasibility of inferring the root language and language lens through training:

Among the many automatic translation systems in use today, it is common practice to gradually build up systematic translation memory and extended translation capabilities through training. But we still haven't seen a translation system with high reliability and high translation accuracy. The training efficiency is likely low due to the use of pure corpus training methods. When the training reaches a certain level, the amount of corpus required to further improve reliability is much higher than people's imagination. That is to say, Although the amount of available corpus is enormous, the amount of corpus required for translation system training is an order of magnitude higher.

However, logically speaking, using the same high-quality translation corpus, it is relatively easy to deduce and confirm from the two languages the root language which is shared by the two languages, and thus the conversion mechanism between languages, namely the language lens, by stripping off the artificial rules of the language. Conceivably, although the existing high-quality translation corpus is still not enough to create an accurate translation system through pure training methods, it is probably enough to infer the root language and language lens.

Summary and Outlook:

When a three- or four-year-old child is brought to his grandparents' home far away from his birthplace, there will be a period of silence at first, but soon he will naturally begin to communicate in a new language, and can rapidly reach a level of proficiency comparable to that of local children. When he returns to his birthplace then, he also only needs a short period of adaptation to regain fluency in his original language. And what's interesting is that he doesn't feel any discomfort caused by this language

switch, and he doesn't even feel the existence of this switch. That's because he created two language filters for his root language that can be switched naturally when needed.

From ancient times to the present, translations have always been done by bilingual people because they know how to express themselves appropriately in the other language (unconsciously converted through a linguistic filter). Later on, machine translation also followed the direct conversion method of the two languages, although it was found in practice that this method has always been unable to meet people's high expectations for translation. The reason is probably that the quality of machine translation is unsatisfactory due to the lack of the bridge of the root language.

Root language, the hidden language of on the level of human thought, which has been proved by many sources, has the simplest and clearest logical structure, without artificial grammatical rules. It contains all the information in natural language, as well as extralingual information that language cannot express directly. It is of great significance for sorting out the relationship between natural language and the thought, and even re-examining the translation ideas and logic that people have always thought to be self-evident.

Through the root language, and by means of advanced filters and meaning assignment, it may be possible for machines to faithfully and accurately convey the meaning between words of a language, to achieve reliable congruent language conversion. As a result, it is expected that the communication barrier between human languages will be eliminated.