



HAL
open science

Understanding: How to Resolve Ambiguity

Shunpeng Zou, Xiaohui Zou

► **To cite this version:**

Shunpeng Zou, Xiaohui Zou. Understanding: How to Resolve Ambiguity. 2nd International Conference on Intelligence Science (ICIS), Oct 2017, Shanghai, China. pp.333-343, 10.1007/978-3-319-68121-4_36 . hal-01820933

HAL Id: hal-01820933

<https://inria.hal.science/hal-01820933>

Submitted on 22 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Understanding: How to Resolve Ambiguity

Shunpeng ZOU¹ and Xiaohui ZOU^{1,2}

¹ China University of Geosciences (Beijing) 29 Xueyuan Road, 100083, China

² SINO-US Searle Research Center, 10 Ronghua Middle Road, 100176, China
949309225@qq.com and zouxiaohui@pku.org.cn

Abstract. This article aims to explore the question: understanding, interpreting, translating, how to resolve ambiguity? Or: how does man-machine combination resolve ambiguity? In order to focus on the essence of the problem, the method is that: target analysis butterfly model and its use cases, macroscopic analysis ambiguity model and its use cases, microscopic analysis matrix model or search model within a series of bi-list and its use cases. The result is through the three examples, from manual translation to machine translation and translation memory on view, pointed out that the fundamental way to resolve ambiguity. Its significance is that the method can be advanced to the generalized translation and corresponding interpretation and final practical understanding, the specific performance is that through man-machine collaboration, and its verifiable results with this method, we can work to resolve various ambiguities better, to ensure accurate understand, prevent and eliminate all kinds of misunderstandings.

Keywords: Linguistic cognition, Mind philosophy, Brain-Machine Integration, Attribute Theory Method

1 Introduction

This article aims to explore the question: understanding, interpreting, translating, how to resolve ambiguity? Or: how does human-machine better resolve ambiguity?

Nature language contains a variety of ambiguities. It is difficult to bypass problem of understanding, interpretation and translation. That is a great challenge for human and computer. We know that “ambiguity in language is an essential part of language, it is often an obstacle to be ignored or a problem to be solved for people to understand each other” .^[1] James Allen introduces the concepts required to build a NL system without losing you in the psycholinguistics, psychology and philosophy of language. ^[2] “Interpretation is a complex practice that requires the interpreter to fully understand, analyze, and process spoken or signed messages. The interpreter, after processing this information, renders the message into another language.” ^[3] “The Language of food: a linguist reads the menu” .^[4]

Not only the difficulty of natural language understanding is ambiguity everywhere, but also generalized language or generalized text (including size character, formula, graphic, table, sound, image, 3D, living) passing information, imparting experience, expressing knowledge (expert knowledge acquisition and its formal representation and

even repetitive reuse) is bound to be due to the various ambiguities (including disparity and disagreement) for both human and machine and their combination to make judgments or decisions, such problems are a major challenge to both human intelligence and artificial intelligence. “We saw earlier how to use ctrl + = to enter natural language input. Now we’re going to talk about how to set up functions that understand natural language. Interpreter is the key to much of this. You tell Interpreter what type of thing you want to get, and it will take any string you provide, and try to interpret it that way.”^[5] “While most of the neural network techniques are easy to apply, sometimes as almost drop-in replacements of the old linear classifiers, there is in many cases a strong barrier of entry.”^[6] “provides a more coherent and integrated framework for performing both bottom-up and top-down knowledge acquisition.”^[7] “Turing himself provides a hint of the answer, noting that his deceptively simple question requires a robust definition of “think”; without one, the question itself is meaningless.” “Indeed, Turing’s original question” “Brilliant as the Turing test is, its popularity has had one pernicious effect: It has reinforced for many people the comfortable illusion that human intelligence is a meaningful measure of intelligence in general. This is understandable: If you can’t define think or intelligence, refer back to human intelligence as the gold standard.”^[8] “Artificial intelligence meets human intelligence”. “Why did it take so long for neural networks to recognize speech and objects in images at human levels?”^[9] “Linguistically, Harmony maximization corresponds to minimization of markedness or structural ill-formedness.”^[10]

2 Method

In order to focus on the essence of the problem, the following examples, from manual translation to machine translation and translation memory of practical point of view, and then advance to general translation and its corresponding interpretation and the final understanding of the basic point of ambiguity resolution (involving targeting analytic butterfly model, macro analytic triangular model and microscopic analytic enumeration matrix model or search model within a series of bi-list query).

2.1 Butterfly model and its use cases

From the "pyramid model" (figure 1a) through the reverse thinking and model reconstruction - that is, the construction of "butterfly model" (figure 1b, it is for us to "understand, explain, translate, how to resolve ambiguity?" What is the role, value and significance of these two models?). We find a paradox existing in the translation pyramid model (figure 1a) .^[11] From artificial and machine translation "pyramid model" (figure 1a) to bilingual "butterfly model" (figure 1b) , there are differences and links between the two basic assumptions involved in the conversion principle. The former is the father of machine translation Weaver (interlanguage hypothesis) and computational linguistics of the international master B. Vauquois (translation pyramid) combination, the latter is the younger generation Zou Xiaohui's unintentional discovery between the two sides both in the actual process of natural language understanding and

in establishment of bilingual coprocessing system ManMachine interaction. To this end, published a corresponding articles on the model was introduced, the focus of which also led to the academic director Professor Feng Zhiwei's concern and the occurrence of the corresponding academic debate and exchange of ideas, and now, not only this academic idea exchange has been effectively promoted, but also the (the narrow sense, the generalized, the alternative) three kinds of bilingual co - processing system, and further discovered its implied the three basic laws (principles or rules) in logic, mathematics and linguistics.^[12]

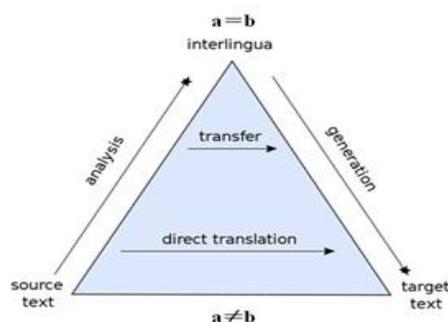


Figure 1a



Figure 1b

Figure 1 Butterfly Model (pyramid model to remove the spire after the reverse reconstruction of the reverse thinking results)

You can see that from artificial and machine translation "pyramid model" (figure 1a) to bilingual "butterfly model" (figure 1b), there are differences and links between the two basic assumptions involved in the conversion principle.

Targeted analysis of butterfly model is characterized by first explained each kind of language both in bilingual model internally, and then translated from each other.

The specific performance of wisdom is for value-taken and confidence-building.

Through using basic methods of value-taking and confidence-building of language by the author's fundamental research in general linguistics taking Chinese language as an example, particularly in exploring how to solve the sharp conflict between two schools of scholars among workers, who adhere to Chinese character based linguist theory and word based linguist theory respectively. Specifically, the purpose of this paper is to solve the conflict between both Chinese character based view and word based view by using the basic unit of Chinese theory.^[13]

Identification, distinguishing and selection of lexical ambiguity are achieved by the selection of term or value-taken. Structural ambiguity distinguishes the function of the phrase and the sentence in their context. English morphemes and Chinese characters and their internal radicals, English words and Chinese words, English phrases and sentences and Chinese corresponding structural units, are almost all difficult to recommend a corresponding function of the relationship. Its roots lies in characters, not only the existence of the word information (its internal radicals), and the existence of inter-word information and even outside the word information. These three aspects of the value-taken and confidence-building on the two aspects, although the system can be used to do a series of processing, but a series of values, a series of confidence, how these two aspects of specific match? Human, computer, bi-brain, between the three, their operation and algorithm and the combination of the model there are differences. So, natural language understanding and how does resolve ambiguity? Be sure to follow the basic program given by the Butterfly Model, ie, (in the respective native language system, respectively) to explain them, (and then in both the phrase and the sentence) to translate them. Thus, understanding, interpreting, translating, three intellectual activities, incompetence is mental intelligence activity, or artificial intelligence activities, and even man-machine bi-brain wisdom, can be divided into two basic activities: the value-taken (select each match) and the confidence-building (sure that the common choice of the selected item is matched).

The following is illustrated and demonstrated in connection with the embodiments:

In Chinese, the relationship between characters and their strings as partial word is understood as formal structure between the object language and the explanatory or interpretation language or meta - language. At the same time, in English, the formal structure is understood by the relationship of words, phrases and sentences.

Figure 1b covers: lexical ambiguity (1) and structural ambiguity (2) which divided into two levels: phrase (ZiZu1) and sentence (ZiZu2). As a result "first explanation" and "re-translation" of the principles, therefore, the above two types of ambiguity and confidence-building process, resolved in the mother tongue groups. In other words, the translation is based on the bilingual expression of knowledge ontology. That is, unambiguous text conversion between source language and target language (which means that bilingual formal expressions refer to a specific object or concept).

For example, the performances of physics, meaning, grammar in English are three words, but three phrases (ZiZu1) in Chinese. Therefore, from the level of Chinese characters to the level of English words, there is no way one by one correspondence. In other words, English words and Chinese words are not entirely at one level.

English and Chinese phrases (ZiZu1) and sentences (ZiZu2) these two levels can be in the same level.^[14]

How does this ambiguity resolve? We use butterfly model to try the list explained as follows: Table 1 with large string (of character) formula to interpret Chinese characters and their relationship between characters, thus, resolve ambiguity.

Table 1. Resolve ambiguity with string of character formula.

String (Character) Formula	Example of Word	String of Character
----------------------------	-----------------	---------------------

物+理=物理 (Material + Law = Physical)	Physical 物理	= Material + Law 物质 (运动的) 基本规律
意+义=意义 (Idea + Principle = Meaning)	Meaning 意义	= Idea + Principle 意念 (确立的) 原则含义
文+法=文法 (Text + Rule = Grammar)	Grammar 文法	= Text + Rule 文本 (组合的) 基本法则
$x + y = z$	Z	$= x + y$

The equation of string (character) formula 1.

$$x + y = z \tag{1}$$

2.2 Triangular pyramid (tetrahedron) model and its use cases

From the "semantic triangle" and "physics, meaning, grammar" (that covering "grammar, semantics, pragmatics") through the combination of Chinese and Western create three-dimensional Triangular pyramid (tetrahedron) model (the most ambitious knowledge system frame and the most fundamental knowledge classification).

What is the use, value, and significance of that? Understanding, explaining, translating: How better to resolve ambiguity by using the models?

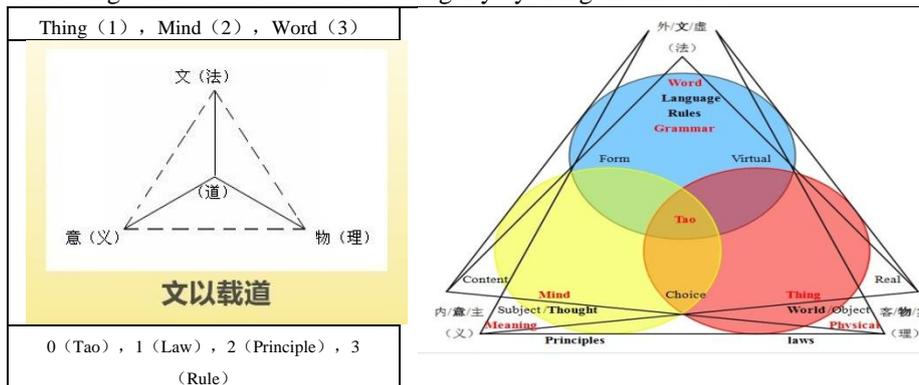


Figure 2a

Figure 2b

Figure 2 Triangular Pyramid (Tetrahedron in Figure 2a) Model is Macro model. It is a broad text (文), Tao is basic true information (covering all phenomenon information).

Combining with Table 1, you can understand the four fundamental categories and the six basic concepts both in Figure 2a and Figure 2b. The four fundamental categories: physical world (real object), meaning thought (subject content), grammar language (virtual form), Tao (laws or principles or rules); the six basic concepts: object and subject, content and form, virtual and real.

First of all, the three composite concept of decomposition, in the Triangle model will be a character set of Chinese visualization of the intuitive advantages, specifically expressed as three sets of character string formula (See: Table 2).

In the fundamental categories (See: Figure 2 and Table 3), introduction of three pairs of concepts: object and subject, content and form, virtual and real, highlight the scope of the phenomenon (world/thing, thought/mind, language/word).

Chinese category Tao is fourth fundamental category (See: Figure 2), the basic concept different from these Western concepts "physics, meaning, grammar" and the two sides behind the outlook. The theory of "holism" and "reduction theory" are not clear (the biggest ambiguity problem is highlighted). Therefore, the categories of thing, mind, word and the essence (law, principle and rule) can be further Explored.

The fundamental category contains more basic categories, which is composed of two groups of three pairs of concepts, they are: object and subject, content and form, virtual and real. The following is illustrated in connection with the embodiments:

Table 2. Resolve ambiguity by using the model in Figure 2a and Figure 2b.

Phenomenon	Essential Nature	Resolve ambiguity
(物理的)物: 世界 (Physical) World	Law (物之)理	Law of material movement 物质运动的规律
(意义的)意: 思想 (Meaningful) Thought	Principle (意之)义	Principle of idea / intention 意念生成的原则
(文法的)文: 语言 (Grammatical) Language	Rule (文之)法	Rules of text construction 文本建构的法则

道Tao

2.3 Query model and its use case

Figure 3b

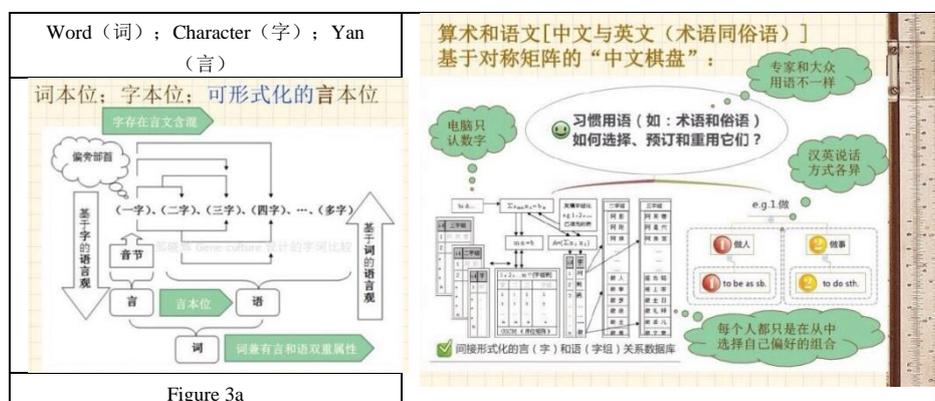


Figure 3a

Figure 3 Query Model (Three kinds of views in linguistics and three bilingual division methods for that from translation to query, laid the basis for both theoretical analysis and practical operation.)^[15]

As it can be seen from Figure 3a, from English word and Chinese character, we extract bilingual Yan and Yu, for all kinds of object language and meta-language that is, out of semantic language, laid a solid foundation for formal understanding.^[18]

It can be seen from Figure 3b that the relational data structure within generalized bilingual object language and meta-language, man-machine cooperation to complete

both Chinese information processing (find the best algorithm) and natural language understanding (to resolve ambiguity), laid solid basis by using two categories formal strategy. It is a actually global (language) positioning system (GLPS) .

Three types of bilingual co-processing with multiple interfaces and can reflect a series of bilingual conversion, not only to answer "understanding, interpretation, translation, how to resolve ambiguity?" for us, but also the problem has a practical effect, value and significance, moreover, it can also be seen as a test "how to better resolve ambiguity with man-machine combination?" Bi-brain and bi-intelligence collaboration system is intuitive prototype.^{[16][17]}

From "monolingual model" and "multilingual model", we abstract "bilingual model" (especially the "three kinds of bilingual co-processing" and its implied logical sequence and position, function linkage, translation promotion or explanation expansion. Resolve ambiguity and understand language.

Human-Machine Twin System as Collaboration Translation

Based on the three basic laws of human-computer cooperation (three basic rules of logical sequence-position, function linkage, translation promotion) and constructed a series of bi-list, using computer-aided, not only can achieve the joint operation between arithmetic and language, but also to the joint operation combined fully automatic, semi-automatic, big collaboration, by the narrow language and its plain text of the string, through the promotion of characters such as the intermediary model, to further promote to generalized language or text (including characters, formulas, graphics, tables, sounds, images, 3D, living).

The following is illustrated and demonstrated in connection with the embodiments:

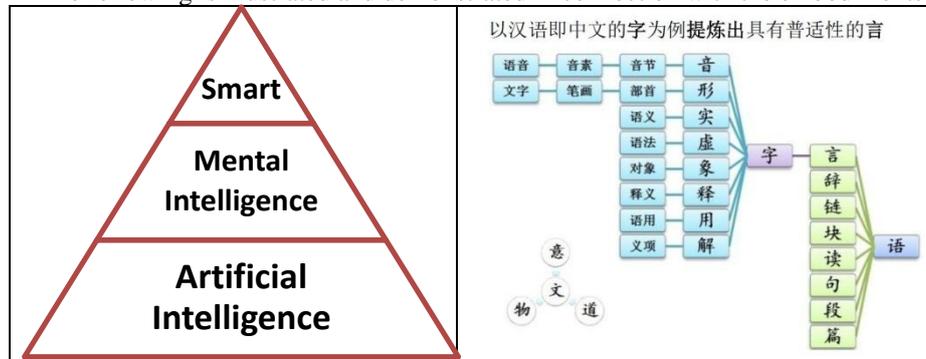


Figure 4a

Figure 4b

It can be seen in Figure 4a that smart (city), mental intelligence and artificial intelligence (different from ancient Greek wisdom in philosophy), converge on the connotation and extension of research object of intelligent science.

As it can be seen in Figure 4b, Chinese as a typical natural language, we understand its own and its unambiguous formal expression, such as combination in Table 3, we can do further research and discussion. Later, in conjunction with Figure 5a, we can also classify it.

Sequence and location of syllables, select: value-taken and confidence-building.

The so-called value-taken, here is that: around a specific goal, given a series of Chinese characters with appropriate structures. For example, in the following table 3, two groups have been accurately displayed, and their features are clearly displayed.

Table 3. Characters=Yan, Chinese=Yu, word and sentence have attributes of both Yan and Yu.

Yan: Monosyllabic character (Zi) (Yan + Yu = Word or Sentence)	Example:	Yu: Two-Syllable and Multi-Syllable Phrase or Sentence (ZiZu1or2)
(音)字(Phonological) Yan		(Linguistic) Yu语 (言)
(形)字(Morphological) Yan		(Semantic +Lexical) Yu语 (辞)
(象)字(Object Language) Yan		(Grammatical) Yu语 (链)
(释)字(Meta-Language) Yan		(Semantic +Grammatical) Yu语 (块)
(实)字(Semantic) Yan		(Pragmatic +Classical) Yu语 (读)
(虚)字(Grammatical) Yan		(Pragmatic +Vernacular) Yu语 (句)
(用)字(Pragmatic) Yan		(Pragmatic +Rhetoric) Yu语 (段)
(解)字(Lexical) Yan		(Pragmatic +Article) Yu语 (篇)

The equation of string (character) formula 2.

$$Yan (Zi, Sentence) + Yu (ZiZu1, Phrase; ZiZu2, Sentence) = Word or Sentence \quad (2)$$

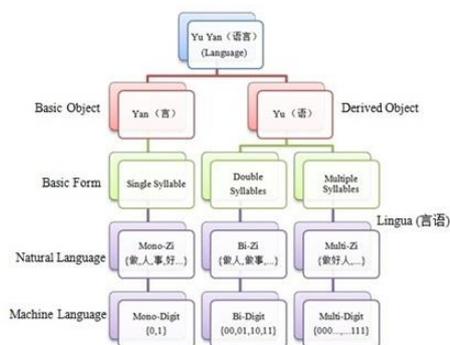


Figure 5a

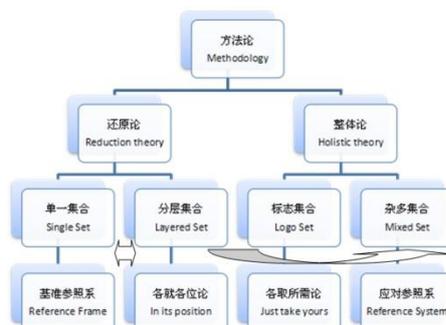


Figure 5b

It can be seen in Figure 5a that the linguistic view introduced here is essentially a broad bilingual view. Between formal language shown in Figure 5a and formal system developed in Figure 5b by the part of reduction theory, its example highly consistent. In other words, the formal understanding natural language here with the arithmetic language (including: machine language) formal understanding, in the same way. The type calculation of the hierarchical set is combined with the attribute set, and the bridge between the simplest single set and the most complex cluster is built up. Thus, language theory, epistemology, object theory, can be unified into the overall logic of sequence and location in twin matrix. Pure formal various functions and a variety of property functions, between each other can be in the agreed under a variety of specific constraints, the formation of the corresponding linkage function. This is the general translation, paved the way. Furthermore, not only for various forms of class functions, but also for various attribute functions (involving the content) can be done: both can be calculated, but also can statistic and can do similar measures.^[19]

3 Conclusion

If Global Positioning System(GPS) can help people to accurately determine the starting point and the target point and the optimal path between each other, then the language, knowledge concept, software object Global Positioning System also can help people clearly understand (interpret, translate or convert) them. In this study, the macroscopic triangular pyramid (tetrahedron) model and the microscopic table model are combined to form Global (Language, Knowledge, Software) Positioning System, and the target (language, knowledge, software) positioning effect. For example, the (eight) Yan and Yu, (three)Zhi (from wisdom to smart, mental intelligence and artificial intelligence), all can be particularly understood. Another example is the further understanding of physics, meaning and grammar (divided into three groups). With this basis, the other types of subdivision and rough (language, knowledge, software) target point of sequencing positioning can be accurate and correct. If both Figure 4a and Figure 2 are to look at the problem from a macro perspective, then Figure 4b and Figure 1b all look at the problem in the middle of the view, while Figure 3b looks at the problem in a microscopic point of view. So that the combination of macro and micro to do natural language understanding and formalized the expression of knowledge, can be accurate, of course, naturally can be resolved which contains a variety of ambiguities.

3.1 Result

The result is through the three groups of examples, from manual translation to machine translation and translation memory on practical view, pointed out that the fundamental way to resolve ambiguity.

The three models and their used cases tell us that each item of character or word as value-taken can be recorded, not only by computer-assisted enumeration into an electronic dictionary, but also by an enumerable corpus at all levels, in particularly through a large number of bilingual expressive knowledge bases that are precisely processed and validated for bilingual linkage. This has laid a large data environment

for repeated application within the three models. Their function can be used by the majority of teachers and students to the extreme. Thus creating a basic development platform can be built for knowledge production.

3.2 Significance

Its significance is that the method can be advanced to the generalized translation and corresponding interpretation and the final practical understanding, the specific performance is that we can work to resolve various ambiguities, to ensure accurate understand, prevent and eliminate all kinds of misunderstandings, through human-computer collaboration with this method and its specific verifiable results.

The practical significance of this paper is not only that it demonstrates how a series of ambiguities between Zi (字) and Zhi (智) are gradually resolved by the three models of macro and micro (and its practical application methods and tools), but also in the process of argumentation. The basic views and basic approaches adopted have universal roles, values and meanings. For the author's innovative ideas and innovative methods of the system argument, has been beyond the scope of this article, interested in it, can refer to the author's related works.

Do they all together can be classified as Tao trying to characterize the object form and the subject content or idea? (To be further explored)

References

1. Cecilia Quiroga-Clare. Language Ambiguity: A Curse and a Blessing. *Translation Journal*
2. James Allen. *Natural Language Understanding*.
3. *Understanding Interpretation and Translation*.
4. Dan Jurafsky. *The Language of Food: A Linguist Reads the Menu*. October 13, 2015
5. *Natural Language Understanding*.
6. Goldberg, Yoav. A Primer on Neural Network Models for Natural Language Processing. *Journal of Artificial Intelligence Research* 57 (2016) 345–420
7. Enrico Motta, Tim Rajan, and Marc Eisenstadt. A Methodology and Tool for Knowledge Acquisition.
8. Brad Allenby. Why it's a mistake to compare A.I. with human intelligence?
9. Terry Sejnowski; Francis Crick Chair. Deep learning: Artificial intelligence meets human intelligence.
10. Paul Smolensky. *Harmony in Linguistic Cognition*.
11. Zou Xiaohui. paradox existing in the translation pyramid model. *X Mind*. Aug 01, 2011
12. Zou Xiaohui. From translation pyramid model to bilingual butterfly model.
13. Zou Xiaohui. Value-Taking and Confidence-Building of Language. AAAS Annual Meeting
14. 邹晓辉, 邹顺鹏. 当代中国大学新使命: 基于汉语思维与双语处理的文化传承和创新 [J]. *南京理工大学学报(社会科学版)*. 2012(05)
15. 邹晓辉, 邹顺鹏. 言和语的关系——在语言和言语的区分基础上进一步 [A]. 第四届中西语言哲学国际研讨会论文摘要集 [C]. 2012
16. 邹晓辉, 邹顺鹏. 两大类形式化方略 [J]. *计算机应用与软件*. 2013(09)
17. 邹晓辉, 邹顺鹏. 双语信息处理方法及原理 [J]. *计算机应用与软件*. 2015(11)

18. Xiaohui Zou.FUNDAMENTAL LAW OF INFORMATION: PROVED BY DOUBLE MATRICES ON NUMBERS AND CHARACTERS. AAAS Annual Meeting 2017
19. Xiaohui Zou.Characteristics of Information and Its Scientific Research. IS4SI-2017 » Session Conference FIS 2017: The Seventh International Conference on the Foundations of Information Science
20. LNCS Homepage, <http://www.springer.com/lncs>, last accessed 2016/11/21.