Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 2962–2969
Marseille, 11–16 May 2020

Construct a Sense-Frame Aligned Predicate Lexicon for Chinese AMR Corpus

Li Song[1,2], Yuling Dai[2], Yihuan Liu[2], Bin Li[2], Weiguang Qu[3]
1. Department of Chinese Language and Literature, Tsinghua University, Beijing 100084, China
2. School of Chinese Language and Literature, Nanjing Normal University, Nanjing 210097, China
3. School of Computer Science and Technology, Nanjing Normal University, Nanjing 210023, China
songli1105@sina.com

Abstract

The study of predicate frame is an important topic for semantic analysis. Abstract Meaning Representation (AMR) is an emerging graph based semantic representation of a sentence. Since core semantic roles defined in the predicate lexicon compose the backbone in an AMR graph, the construction of the lexicon becomes the key issue. The existing lexicons blur senses and frames of predicates, which needs to be refined to meet the tasks like word sense disambiguation and event extraction. This paper introduces the on-going project on constructing a novel predicate lexicon for Chinese AMR corpus. The new lexicon includes 14,389 senses and 10,800 frames of 8,470 words. As some senses can be aligned to more than one frame, and vice versa, we found the alignment between senses is not just one frame per sense. Explicit analysis is given for multiple aligned relations, which proves the necessity of the proposed lexicon for AMR corpus, and supplies real data for linguistic theoretical studies.

Keywords: predicate frame, semantic role, Abstract Meaning Representation, lexicography

1. Introduction

Sentences are the basic units of language use. Semantic analysis on sentence-level focuses on semantic structures of the sentences. The backbones of semantic structures are the various semantic relations contained in the event frames of predicates, mainly the dominating-dominated relations between predicates and nominal components. Therefore, the study of semantic role frame of predicates is of great importance to semantic analysis.

However, the way of defining semantic roles has always been controversial in the linguistic field. It is unavoidable to explore methods to describe semantic relations and determine the granularity of semantic role labels. Xue (2006a) pointed out that existing annotated resources define semantic roles through different levels of abstraction. At present, there are three general methods for defining, namely, predicate-general, frame-specific and predicate-specific, but none of them is perfect. Predicate-general labels are not suitable for representing core semantic roles that bear many dynamic problems. The frame-specific method removes the limit to the number of labels while causes complex labels and burdens annotators. The predicate-specific method can effectively annotate the dynamic core semantic roles, but is hard to handle non-core semantic roles. (See Section 2 for details.) Therefore, we argue that it is more appropriate to use predicate-specific labels for core semantic roles and predicate-general labels for non-core semantic roles. English Proposition Bank (PropBank) (Palmer et al., 2005) and the emerging sentence semantic representation method, Abstract Meaning Representation (AMR) (Banarescu et al., 2013), both adopt such a method. Core semantic role labels used by the two are arg0 - arg4, but non-core semantic labels differ greatly in granularity. PropBank uses 13 non-core semantic role labels with coarse granularity and low discrimination, while that number of AMR is up to 40, and Chinese AMR (CAMR)[1] (Li et al., 2016) adds another 4 to adapt to the characteristics of Chinese, showing relatively fine granularity and appropriate discrimination. Besides, AMR allows the addition of omitted semantic roles, such as the head "人 (person)" in "受伤的 (the injured)", which helps a full understanding of sentence meaning. Thus, in general, AMR has a greater advantage in semantic role annotation.

Since core semantic roles are represented by predicate-specific labels, the construction of AMR corpus is inseparable from the support of predicate frame lexicon. The lexicon currently used by CAMR is extracted from the annotated corpus of Chinese Proposition Bank (CPB) (Xue and Palmer, 2009). It contains 26,650 core semantic role frames of 24,510 Chinese predicates under different senses. Since one sense constitutes one frame in the CPB lexicon, each predicate (word token) is numbered according to the order of senses there. Figure 1 shows an example of how CAMR uses the lexicon to annotate semantic relations between a predicate and its core semantic roles. "还-01" refers to the first sense "give back" of "还", which has the other sense "counteroffer", in the CPB lexicon, and it has three core semantic roles, *arg0* "returner", *arg1* "thing returned" and *arg2* "returning to", which correspond to "我", "书" and "他" in the sentence respectively.

我[1] 把[2] 书[3] 还[4] 给[5] 他[6] 。[7]
I BA book give back to he .
I gave the book back to him.

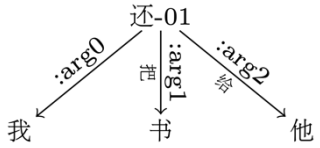

Figure 1: An Example of CAMR

However, when annotating CAMR corpus, we realize that senses and frames are not in one-to-one correspondence. A frame may align with multiple senses and a sense with multiple frames. On the one hand, a frame can align with different senses. For example, "打 (beat)" has 26 senses[2] that all have core semantic roles, 20 of which align with a same frame which means someone uses his hand or other tools to perform an action on something. On the other hand, a sense

[1] https://catalog.ldc.upenn.edu/LDC2019T07.

[2] *Modern Chinese Dictionary* (7th ed.) contains 24 senses of "打" such as "hit", "knock", "pack", and the other two senses are supplemented based on authentic language materials.

Definition Method	Resource	Language	Distinction between Core and Non-core Labels	Amount of Role Labels
Predicate-general	VerbNet	English	No	36
	Sinica Treebank	Traditional Chinese	Yes	60 (12+43) 5 more noun labels
	Chinese NetBank	Simplified Chinese	Yes	22 (10+12)
Frame-specific	FrameNet	English	Yes	1224 frames
	CFN	Chinese	Yes	323 frames
Core Role: Predicate-specific Non-core Role: Predicate-general	PropBank	English	Yes	18 (5+13)
	CPB	Chinese	Yes	18 (5+13)
	AMR	English	Yes	45 (5+40)
	CAMR	Chinese	Yes	49 (5+44)

Table 1: Semantic Role Labels for Typical Semantic Role Annotated Resources in English and Chinese

can align with different frames. For example, the monosemous word "安排 (arrange)" can correspond to three different frames, including "fix someone up with something", "assign someone to do something" and "put someone somewhere". Therefore, we consider that it is not reasonable to blur senses of predicates with core semantic role frames. In addition, data shows that the one-to-one correspondence of the CPB lexicon makes it difficult to improve the quality of annotation and the result of automatic analysis of CAMR. Therefore, we decide to reconstruct a predicate frame lexicon which has a double-level numbering to index frames and senses. The new lexicon can match CAMR annotation scheme better.

The rest of this paper is organized as follows. Related work on predicate frames is referred to in Section 2. In Section 3, we introduce how to number and describe senses and frames in the new lexicon. In Section 4, we discuss diverse alignment cases between senses and frames based on data and reasons behind them. Section 5 classifies types of a sense aligning with multiple frames. Conclusions and future work can be found in Section 6.

2. Related Work

Many linguists are striving for the research on predicate frame with theories, including Valence Theory (Tesnière, 1959) based on verbs, Case Grammar (Fillmore, 1977a; 1977b) based on nominal components and Frame Semantics (Fillmore, 1982) based on events. And many resources have been constructed according to these theories such as lexicons, framesets and annotated corpora. Generally, there are three methods for defining semantic roles at present, namely, predicate-general, frame-specific and predicate-specific, but none of them is perfect.

The predicate-general method uses predicate-general labels such as "agent", "cause" and "tool" to annotate semantic elements. These labels can be used to all predicates. Generally, the label set adopted this method contains about 50 labels and shows a moderate granularity, not very fine but fine enough for their purposes. Therefore, there is no need to construct a predicate frame lexicon, but these labels fail to flexibly annotate core roles. For example, an element may serve as more than one role. In the sentence "She eased my pain", "she" functions as an "agent" and a "cause" at the same time, which cannot be completely annotated by predicate-general labels. Typical resource banks constructed in this mothod are VerbNet (Kipper et al., 2000; Kipper et al., 2004; Kipper et al., 2006), Sinica Treebank (Chen et al., 2003) and Chinese NetBank of Peking University (Yuan, 2007).

The frame-specific method uses labels that obtain meanings from specific events to annotate semantic roles. Different frames are assigned different element labels, which means the number of labels is unlimited. Therefore, it leads to fine granularity and huge number of labels, burdening annotators and parsers. Typical resources are FrameNet (Baker et al., 1998), based on the theory of Frame Semantics, and FrameNets for other languages, including ASFALDA (French FrameNet), CFN (Chinese FrameNet), FrameNet Brasil (Brazilian Portuguese FrameNet), SALSA (German FrameNet), SFN (Spanish FrameNet), etc.

The predicate-specific method annotates semantic roles with labels that are only meaningful to specific predicates (such as arg$x$ ($x \in$ N)). Each predicate (more specifically, each sense of each predicate) has a set of specific semantic roles, facilitating the annotation of core semantic roles. In light of this, a static predicate frame lexicon must be constructed. But it is not suitable to represent non-core semantic roles, because semantic relations of time, place, reason, etc., are universal to all predicates. Resources adopting this method include PropBank and NomBank (Meyers et al., 2004) based on Penn TreeBank. And there are other versions of different languages such as Urdu, Chinese, Arabic and Finnish PropBank, and Chinese NomBank (Xue, 2006b). It is noted that core semantic role labels used by AMR and CAMR are the same as PropBank.

PropBank specifies 5 predicate-specific labels (arg0 - arg4) to establish core semantic role frame for each sense of each predicate. For non-core semantic roles, PropBank uses 13 predicate-general labels for all predicates, including "LOC (location)", "DIR (direction)", "CND (condition)" and so forth. CPB (Xue and Palmer, 2009) follows the PropBank system and utilizes the same 5 core semantic role labels that are predicate-specific and 13 non-core labels that are predicate-general.

AMR, a new sentence semantic representation method, and CAMR, a Chinese version which inherits its scheme, both adopt this method. As to core semantic roles, AMR, CAMR and PropBank set the same 5 predicate-specific labels that are only meaningful to specific predicates. AMR sets 40 non-core semantic roles labels to get a finer granularity, while CAMR adds 4 more to satisfy the need of Chinese annotation, totaling 44. Table 1 lists basic information of typical English and Chinese relevant resources.

Word	Sense/Frame ID	Frame
挨	01	arg0: *endurer*; arg1: *stuff endured*
挨	02	arg0: *agent*; arg1: *entity arg0 is close to*
拨弄	01	arg0: *agent*; arg1: *theme*
拨弄	02	arg0: *agent*; arg1: *theme*
拨弄	03	arg0: *agent*; arg1: *theme*
替代	01	arg0: *replacement*; arg1: *entity replaced*
替代	02	arg0: *replacement*; arg1: *entity replaced*; arg2: *agent*

Table 2: Three Word-examples in the CPB Lexicon

Word	POS	Pinyin	Sense	Sense ID	Frame ID	Arg0	Arg1
鸣	v	míng	（鸟兽或昆虫）叫 ((birds, beasts or insects) chirp)	1	1	*agent*	*thing arg0 makes sound by*
鸣	v	míng	发出声音；使发出声音 (make a sound; to make a sound)	2	1	*agent*	*thing arg0 makes sound by*
鸣	v	míng	发出声音；使发出声音 (make a sound; to make a sound)	2	2	*agent*	
鸣	v	míng	表达；发表（情感、意见、主张） (express; voice (feelings, suggestions, opinions))	3	3	*agent*	*content*

Table 3: A Word-example in the new Lexicon

3. Numbering and Description of Senses and Frames

As mentioned in Section 1, senses and frames of predicates are not one-to-one correspondence as they are in the CPB lexicon, which is currently used by CAMR. Table 2 shows three word-examples in the CPB lexicon. Senses of "挨" are "endure" and "be closed to", which can be distinguished according to the information of frames, while the three senses of "拨弄" are "fiddle with", "manage" and "stir up", which cannot be distinguished. The two frames of "替代" mean "someone takes the place of another" and "someone replaces one thing with another", but "替代" has only one sense "replace". It can be seen that meanings of words cannot be correctly expressed if the lexicon provides only information of frames but neglects information of senses. Therefore, we decide to reconstruct a new predicate lexicon which provides information of both senses and frames for CAMR. To avoid confusions over indexes, we design a double-layer numbering in the new lexicon, which is, numbering senses and frames of each word (word token) respectively to make them interrelated and independent.

We select the latest edition (7th edition) of *Modern Chinese Dictionary* (Dictionary Compilation Office, Institute of Linguistics, Chinese Academy of Social Sciences, 2016), *MCD* for short, to screen words and their senses[3] for the new lexicon, in which senses can be independently used as a word and have core semantic roles, and we retain some useful information in *MCD* such as pinyin, parts of speech and sense explanations, which are very helpful for CAMR annotators to choose the right sense. A word-example "鸣 (cry; sound; express)" of our new lexicon is showed in Table 3 ("POS" is short for "part of speech"), and how to number and describe senses and frames is illustrated later.

In *MCD*, homographs with different pronunciations and some with different meanings are listed separately. Strictly speaking, they should be regarded as different words. However, CAMR annotation platform displays word information according to word tokens, which means it is impossible to distinguish homographs. Therefore, they are distinct from each other in the new lexicon by indexes of senses. The guiding principle of numbering is to sort senses according to the sequence of senses in *MCD*, using a number from "1" and increasing it consecutively. Besides, one additional number is added before the initial index to differentiate entries of predicates in *MCD* when necessary. The second entry gets "3"[4] before the initial index and the third entry gets "4", with the number increasing in sequence. That is to say, if the second entry has 10 senses, the tenth sense will be numbered as "310". For example, the new lexicon contains two entries of "上(up)". 2 senses are included in the first entry and 13 in the second entry, so indexes of these senses are "1, 2, 31, 32... 39, 310... 313". Indexes of frames, however, are independent of indexes of senses. For each word (word token), frame number increases progressively from "1".

As we can estimate, although the length of sense and frame indexes in the new lexicon can increase infinitely in theory, double-digit frame indexes are and will not appear in the future[5], so a frame index can precede a sense index and its length can be limited to 1. Unlike CPB of adopting equal length numbering, senses are indexed with unequal length numbers in the new lexicon, restricting digits to be 1, 2 or 3. Usually, double digits are enough to number predicates in CAMR with the new lexicon (the first digit is a frame index, and the second digit refers to a sense index, both of whom increase in order from "1"), and the length of a whole index is no more than 4 (the first digit is a frame index, and the second, third and fourth digits constitute a sense index). According to indexes of the lexicon constructed at present, only 5.42% and 0.08% of total data are sense indexes numbered in two or three digits, and average

[3] We will add senses that are not included in *MCD* to the new lexicon when needed.

[4] The added number starts with "3" instead of "2" because an entry may contain more than 20 senses.

[5] None of words with complex meanings in *MCD* has 10 frames.

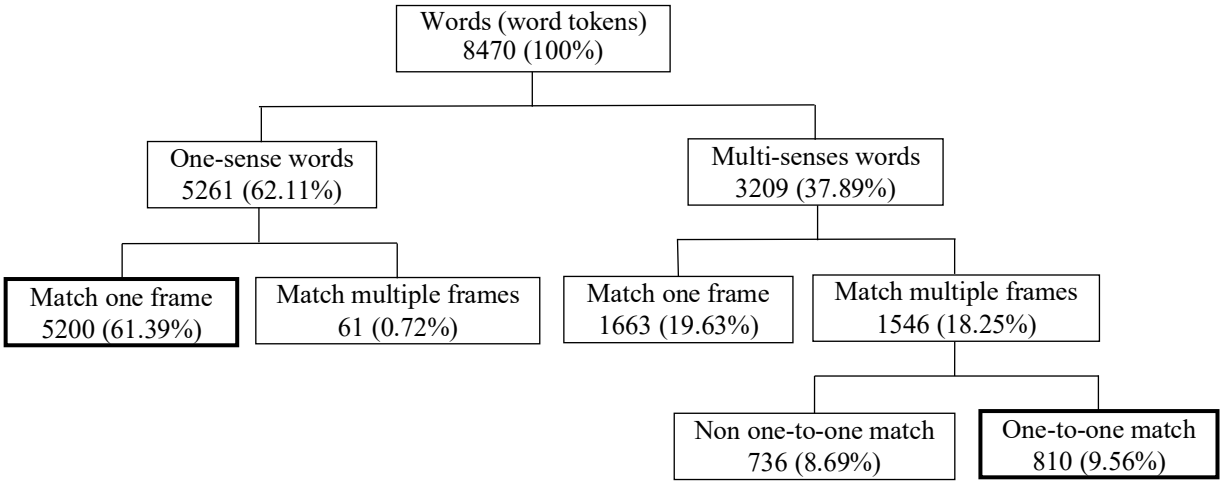

Figure 2: Distribution Map of Senses and Frames in the New lexicon

digits are 2.06. It is basically the same as previous annotation data based on CPB, but the new one can annotate core semantic role frames more accurately. Meanwhile, for words with only one frame, we can infer that their indexes must be "11", so we can also ease the burden of annotators by modifying annotation platform, and replacing them in batches.

As for the description of core semantic role frames, the new lexicon uses the same way as the CPB lexicon, and still uses English to describe core semantic roles to maintain compatibility with AMR dictionaries.

4. Alignment Cases between Senses and Frames

Different senses of one same predicate are relatively independent, and event frames of predicates seem to be different since senses are different. Therefore, predicate frame lexicons like PropBank usually do not distinguish senses from frames. This scheme conforms to our language intuition. As mentioned in Section 3, however, senses are not entirely in one-to-one correspondence with frames. Therefore, we change the current scheme of the CPB lexicon and number senses and frames for each word (word token) separately. At present, the new lexicon has reached a scale of 14,389 senses and 10,800 frames of 8,470 words (word tokens), enough to allow us to quantitatively analyze them. This section gives a detailed introduction to statistical results of alignment cases, and analyzes them with examples. For the sake of brevity, "multi-" are used to modify two or more items. We use "**one-sense word**" to denote a word (word token), of which one sense can form a word independently and has core semantic roles, and "**multi-senses word**" to denote a word (word token) where two or more senses can form a word independently and have core semantic roles. Note that meanings of "one-sense" and "multi-senses" are different from "monosemous" and "polysemous".

4.1 Data of Different Alignment Cases

8,470 words (word tokens) have been annotated in the new lexicon, of which about three fifths are one-sense words and two fifths are multi-senses words. One-sense word basically gets one event frame, because it only has one meaning and describes an event. Multi-senses word, however, can express different meanings and therefore is more complicated--there is not a consistent one-to-one match between senses and frames. Maybe all senses align with one same frame, or some senses align with a same frame while others align with different frames, or even each sense aligns with each frame. The ratio of senses and frames in the new lexicon is about 1.33:1, which obviously demonstrates that senses and frames are not exactly one-to-one correspondence and it is unreasonable to mix them together.

Detailed information is given in Figure 2. Senses are in one-to-one correspondence with frames in only two cases in bold border. One is that one-sense word only has one frame, for example, the frame of "罢工 (strike)" is "arg0: *agent*"; the other is that each sense gets its own frame for multi-senses word. For example, there are three senses of "等 (wait)": "s1: same in degree or quantity", "s2: use a small steelyard to weigh things" and "s31: wait"[6]. S2 activates a new frame although s2 and s1 are cognates. As for s31, it is just a homograph with s1 and s2. Therefore, each of them gets their own unique frame. In the new lexicon, 70.92% words show a one-to-one correspondence between senses and frames. When only multi-senses word is considered, only 25.24% words meet this requirement, accounting for a quarter of all multi-senses words. This fully proves that it is implausible to blur senses and frames of predicates. Given the great number of instances that senses are not in one-to-one correspondence with frames in Chinese, it is worth researching.

In addition, we observe predicate frame lexicons of PropBank in English, Spanish, Irish and other languages, and

[6] "S1" refers to the sense whose index is "1" and senses in this paper are translated from *MCD*. See Section 3 for details about the method of numbering senses.

Word	Sense/Frame ID	Frame
robar	01	arg0: *thief, agent*; arg1: *thing stolen*; arg2: *stolen from*
modificar	01	arg0: *causer of transformation*; arg1: *thing changing*; arg1: *end state*; arg1: *start state*
modificar	02	arg0: *causer of transformation*; arg1: *thing changing*; arg1: *end state*; arg1: *start state*
llegar	01	arg0: *entity in motion*; arg1: *extent*; arg1: *start point*; arg1: *end point*
llegar	02	arg0: *traveler*; arg1: *destination*
llegar	03	arg0: *entity in motion*; arg1: *extent*; arg1: *start point*; arg1: *end point*
confiar	01	arg0: *truster*; arg1: *thing trusted*
confiar	02	arg0: *relier, needer*; arg1: *thing needed, relied on*; arg2: *for, in order to*
confiar	03	arg0: *entity giving up control, agent*; arg1: *entity entrusted*; arg2: *entity entrusted to*

Table 4: Four Word-examples in Spanish PropBank Lexicon

find that without exception, these languages have different matches between senses and frames. This finding should be applied not only to Chinese, but to many other languages. Let's take Spanish[7] as an example. (1) One-sense word gets a frame. For example, "robar (steal)" is a one-sense word meaning "steal" and it aligns with a frame indicating "someone steals something from somewhere." (2) Senses of multi-senses word align with a same frame. For instance, "modificar (modify)" has two senses: "change, transform" and "modify", both of which align with one frame, indicating that "someone changes something from one state to another". (3) Senses of multi-senses word align with different frames, while their relations are not one-to-one correspondence. For example, "llegar (reach)" has three senses: "move", "arrive on time, succeed" and "come". Both the first and third sense mean "something goes from one place to another through a certain path", emphasizing movement, so they align with a same frame. The second sense means "someone arrives at somewhere/some state" and only emphasizes the result, so it aligns with a frame that is different from another two senses. (4) Each sense of multi-senses word aligns with a unique frame. For example, "confiar (trust)" has three different senses, which are "to trust/to have faith", "to believe in someone" and "to have an expectation". Their corresponding frames are "someone trusts in something", "someone depends on something to do something" and "someone entrusts something to another". Table 4 shows these word-examples in the lexicon of Spanish PropBank. It can be observed that senses and frames of Spanish predicates also have complex relations.

4.2 Analysis of Complex Alignment Cases between Senses and Frames

Liu (2015) analyzes the meaning evolution of 20 core words used from pre-Qin Dynasty to the present age by adopting three traditional modes, which are expansion, reduction and transfer, cognitive metaphor and metonymy as well as cognitive event frame. It is found that frames representing individual psychological cognitive model can shed lights on meaning evolution, and meaning evolution would be subject to the frame activated by its original meaning. We also attempt to analyze the evolution of word meanings from the perspective of frames, and explore reasons why senses and frames match so diversely.

Different approaches to word meaning evolution can be seen from its corresponding event frame. Frame elements not only tend to be more and more abstract, but show dynamic changes of increase, disappearance, merger and decomposition during evolution of word meaning. The most common change for frame elements is getting more abstract. "干燥 (dry)" is originally used only to modify a physical object that lacks moisture, and later to describe a dull event. This way of meaning evolution only involves changes of frame elements themselves, and does not activate new frames, so multiple senses generated align with a same frame. According to statistics, there are 5,477 frames for multi-senses words in the new lexicon, of which 2,585 frames align with more than one sense, accounting for 47.20% of the whole data, nearly half. The increase or disappearance of frame elements is common, too. For example, "恢复 (recover)" refers to the change of something into its original form, and then refers to someone carrying on such a change, adding a frame element to indicate its agent. "摆渡 (ferry)" originally refers to people carrying goods across the river by boat, and later also refers to people crossing the river by boat, which means that the frame element of goods disappears. This way of meaning evolution will lead to number fluctuation of frame elements, so it will activate new frames. What's more, there are also cases where frame elements are merged or decomposed. "干涉 (interfere)" originally refers to one's initiative to connect with a targeted matter, and then also refers to the association between two or more matters without any target. The original agent and patient, therefore, are merged into a collective frame element. "隔膜 (estrangement)" originally describes a situation that things are not related to each other. Then it is used to describe that someone isn't aware of something. Original non-directional frame elements are decomposed, and a new element appears to show directional information. Such ways of meaning evolution will also activate new frames.

Event frames of predicates will remain the same when their frame elements only go through a transition from concreteness to abstraction. Once increase, decrease, merger or decomposition of frame elements are involved, new frames will inevitably be activated. Multi-senses words are formed if different meanings produced during the evolution of

[7] https://github.com/System-T/UniversalPropositions/tree/master/UP_Spanish.

Reasons why a Sense Corresponds to Multiple Frames		Amount of senses	Proportion	
Ambiguous criteria for differentiating and merging senses	No distinction between autonomous meaning and causative meaning	126	78.25%	91.82%
	Neglecting the direction of action	5	3.14%	
	Complex interpretation	15	9.43%	
Modifying directional relations (adjectives)		13	8.18%	8.18%
Total		159	100%	

Table 5: Distribution of a Sense Aligning with Multiple Frames

word meaning are retained. That is to say, there are complex traces of evolution among senses of multi-senses words. Some words have not activated new frames, so all senses align with a same frame. "表演 (perform)" has three senses "s1: perform; show plots or skills", "s2: do typical actions" and "s3: act deliberately". They align with a same frame because they all show a meaning that someone shows audiences a certain state. Some words only have partial senses activating new frames. The basic meaning of "出动 (set out)" is "s1: hang out", which evolves into "s3: act". This change doesn't activate new frames because s1 and s3 both mean someone does actions, but when s1 evolves into "s2: dispatch (army)", it adds the frame element of agent and then activates a new frame because s2 means one make another act. Other words activate a new frame for each new sense, showing a one-to-one correspondence between frames and senses. The basic meaning of "斗争 (struggle)" is "s1: contradictory parties conflict with each other". There is only one collective element in the frame. And when it evolves into "s2: the masses fight against enemies or evildoer", this collective element is decomposed into two elements: the attacker and victim, activating a new frame. In addition, when it evolves into "s3: work hard", the attacker disappears and a new frame gets activated. Therefore, perplexed matches between senses and frames of multi-senses words appear.

5. Analysis of a Sense Aligning with Multiple Frames

1.16% of one-sense words can match multiple frames although most of them in the new lexicon match only one. And there are 159 senses corresponding to multiple frames in the new lexicon, accounting for 1.11% of total senses. We extract all senses that align with multiple frames and analyze them one by one. Two main reasons are responsible for the phenomenon that a sense aligns with multiple frames. One is ambiguous criteria for differentiating and merging senses in *MCD*, and the other is that some words are used to denote actions or states with directional relations between two concepts. The former is embodied in three different types. Table 5 shows the distribution of senses corresponding to multiple frames in the new lexicon, with each being illustrated by examples in the followings.

5.1 Ambiguous Criteria for Differentiating and Merging Senses

The annotation of the new lexicon is based on senses classified in *MCD*, but we find that there are different criteria for classifying senses in this dictionary. For similar events that both involve different participators, some words are divided into several senses, while the other only has one sense that matches multiple frames. The ambiguous criteria for differentiating and merging senses are mainly manifested in the following three types.

Firstly, there is no distinction between autonomous meanings and causative meanings. Actions of some words can affect actors themselves, that is, the action executor is the actor. Besides, actions of these words can also affect other matters, which means initiators and executors of actions are respectively acted by different elements. Some of these words regard autonomous meaning and causative meaning as different senses such as "s8" of "败 (defeat)": "fail" and "s9: defeat". However, some of them combine autonomous meaning and causative meaning into one sense, such as "降低-s1 (decline): decline; cut down". Some senses can also represent causative meanings even they seem to show only autonomous meanings, such as "隐藏-s1 (hide): prevent from being seen or discovered" means someone hides himself or something. We distinguish two different frames for senses that blurs autonomous meanings and causative meanings.

In fact, the autonomous and causative meaning of a word share a same event frame given that automation means making actors themselves act. But we hold that they should be divided into separate frames when annotating, mainly because of the following two considerations. For one thing, words that only have autonomous meanings must be forced to get causative meaning if we want to analyze all autonomous meanings as making actors themselves act, so "走 (walk)" means make oneself walk, and "跳舞 (dance)" make oneself dance, which are obviously not in line with language intuition. For the other, it is easier to merge than to decompose when processing data by computers.

Secondly, directions of actions are neglected. Actions denoted by some words are directional. Some of these words in *MCD* gain different senses based on different directions. For example, "借 (borrow or lend)" can mean borrowing with "s1: take and use something that belongs to others for the time being", and can also mean lending with "s2: allow someone to use your matters or money temporally". Other words neglect the directionality of actions and only have one sense, such as "贷-s1 (loan): borrow or lend". Two different frames are annotated for senses that ignore the directionality of actions.

In fact, events described by these two frames are the same despite different directions. For example, two parties compete in an event described by "败 (defeat)", with one party winning and the others losing. But since CAMR stipulates

that arg0 of a predicate is a prototype agent, that is, the actor of an action, senses with different directions should align with different frames.

Thirdly, interpretations of some words are complicated. They can represent many complex but similar events, so definite senses are hard to get. Therefore, *MCD* merges these complex messes into a same sense. "轰 (bang)" is an onomatopoeia originally, which is later used to describe an action that makes this sound: "轰-01: thunder; (artillery) fire; (gunpowder) explode." It means that something makes a booming sound on its own, or someone makes a booming sound when he strikes another one with something. We annotate such words with different frames based on events they actually represent.

The above three kinds of senses that align with multiple frames are all caused by different criteria for differentiating and merging senses in *MCD*. Hu et al. (1982) pointed out that subjectivity is hard to avoid when dictionaries induce word meanings given that their guiding principles are linguistic intuition and special cases. Word meanings that are induced cannot fully conform to the reality, and boundaries among senses are relative unclear because they show a property of continuous transition, which are inherent issues to differentiate and merge senses. We can provide some references for this issue that exists in dictionaries like *MCD* through our new lexicon.

5.2 Modifying Directional Relation

A sense usually aligns with two different frames when modifying directional relations between two concepts. This rule is set for adjective senses. Adjectives are used to modify nominal phrases, some of which can represent directional relations between concepts, such as one's attitude towards something. Relations inevitably involves several concepts. In an event frame, the relation between concepts can be viewed as a frame element, or different concepts themselves be regarded as different frame elements. Therefore, we argue that adjectives that modify directional relations align withtwo different frames. When such an adjective appears in a sentence, the frame that regards concepts and their relation as the whole element is activated if this sentence highlights relation. Otherwise, the frame that treats concepts as different elements is used instead. For example, when "恶狠狠 (vicious)" is used in "态度恶狠狠 (the attitude is vicious)", the corresponding frame is "arg0: thing described", while the frame for "张三对李四恶狠狠 (Zhang San is vicious to Li Si)" is "arg0: person described; arg1: person arg0 is cruel to". Such adjectives include "和蔼 (kind)", "苛刻 (harsh)", "冷漠 (indifferent)" and so forth.

6. Conclusions and Future Task

Considering the confusion between senses and frames of predicates in the CPB lexicon, it would impact the quality of CAMR annotation and automatic analysis. Therefore, a newly predicate frame lexicon suitable for CAMR annotation scheme is reconstructed. We design a double-level numbering to index senses and frames in the new lexicon and 14,389 senses and 10,800 frames of 8,470 words (word tokens) have been included.

It is proved that senses and frames are not one-to-one correspondence through statistical analysis. Firstly, we distinguish different cases for matching senses to frames, and count the number of predicates in each case. It is calculated that for predicates with two or more senses, cases of one-to-one match only account for one quarter of the total. Then, we explore reasons why there are so many match cases: some words are more and more abstract, and changes of increase, disappearance, merger and decomposition of frame elements cause fluctuation of frame amount, which lead to different senses and frames. Two main reasons are summarized focusing on the case that one sense aligns with multiple frames: one is ambiguous criteria for differentiating and merging senses; the other is that some words modify directional relations between two concepts. Holding conclusions above, we expect to further the study of Frame Semantics theory and provide references for differentiating and merging senses in dictionaries such as *MCD*.

In the future, we will continue the construction of our new lexicon, and apply it into CAMR annotation when its size reaches about 25,000 words. A logical and computability description method is also expected to be designed to describe frame elements of predicates.

7. Acknowledgements

We thank the reviewers. This work is partially supported by project 18BYY127 under the National Social Science Foundation of China, project 61772278 under the National Science Foundation of China, and Project Funded by the project for Jiangsu Higher Institutions' Excellent Innovative Team for Philosophy and Social Sciences.

8. Bibliographical References

Baker, C. F., Fillmore, C. J. and Lowe, J. B. (1998). The Berkeley FrameNet Project. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics and International Conference on Computational Linguistics*, 86-90.

Banarescu, L., Bonial, C., Cai, S., et al. (2013). Abstract Meaning Representation for Sembanking. In *Proceedings of the Linguistic Annotation Workshop and Interoperability with Discourse*, 178-186.

Chen, K. J., Luo, C. C., Chang, M. C., et al. (2003). Sinica Treebank: Design Criteria, Representational Issues and Implementation. *Treebanks: Building and Using Parsed Corpora*. Springer Netherlands, 231-248.

Dictionary Compilation Office, Institute of Linguistics, Chinese Academy of Social Sciences. (2016). *Modern Chinese Dictionary (7th ed)*. Beijing: The Commercial Press.

Fillmore, C. J. (1977a). The Case for Case Reopened. *Syntax and Semantics 8: Grammatical Relations*. New York: Academic Press, 59-81.

Fillmore, C. J. (1977b). Topics in Lexical Semantics. *Current Issues in Linguistic Theory.* Bloomington: Indiana University Press, 76-138.

Fillmore, C. J. (1982). Frame Semantics. *Linguistics in the Morning Calm*, 111-137.

Hu, M. (1982). *Manual of Lexicography*. Beijing: China Renmin University Press.

Kipper, K., Dang, H. T. and Palmer, M. (2000). Class-Based Construction of a Verb Lexicon. In *Proceedings of the National Conference on Artificial Intelligence and Conference on Innovative Applications of Artificial Intelligence*, 691-696.

Kipper, K., Korhonen, A., Ryant, N., et al. (2006). Extending VerbNet with Novel Verb Classes. In *Proceedings of* the *International Conference on Language Resources and Evaluation*, 1027-1032.

Kipper, K., Snyder, B. and Palmer, M. (2004). Extending a Verb-Lexicon Using a Semantically Annotated Corpus. In *Proceedings of* the *International Conference on Language Resources and Evaluation*, 1557-1560.

Li, B., Wen, Y., Bu, L., et al. (2016). Annotating the Little Prince with Chinese AMRs. In *Proceedings of the Linguistic Annotation Workshop at Annual Meeting of the Association for Computational Linguistics*, 7-15.

Liu, X. (2015). *A Study on the Evolution of Word Meaning Based on the Great Chinese Dictionary*. Nanjing Normal University.

Meyers, A., Reeves, R., Macleod, C., et al. (2004). The NomBank Project: An Interim Report. In *Proceedings of the Frontiers in Corpus Annotation Workshop at Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies*, 24-31.

Palmer, M., Gildea, D. and Kingsbury, P. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1): 71-106.

Tesnière, L. (1959). *Eléments de Syntaxe Structurale*. Paris: Klincksieck.

Xue, N. (2006a). A Chinese Semantic Lexicon of Senses and Roles. *Language Resources and Evaluation*, 40(3-4): 395-403.

Xue, N. (2006b). Annotating the Predicate-Argument Structure of Chinese Nominalizations. In *Proceedings of the International Conference on Language Resources and Evaluation*, 1382-1387.

Xue, N. and Palmer, M. (2009). Adding Semantic Roles to the Chinese Treebank. *Natural Language Engineering*, 15(1): 143-172.

Yuan, Y. (2007). The Fine Level of Semantic Role and Its Application in Information Processing. *Journal of Chinese Information Processing*, 21(4): 10-20.