Features of Verb Complements in Co-composition:
A case study of Chinese baking verb using Weibo corpus

Yu-Yun Chang
Graduate Institute of Linguistics,
National Taiwan University
No. 1, Sec. 4, Roosevelt Rd.,
Taipei, Taiwan (R.O.C.) 10617
`yuyun.unita@gmail.com`

Shu-Kai Hsieh
Graduate Institute of Linguistics,
National Taiwan University
No. 1, Sec. 4, Roosevelt Rd.,
Taipei, Taiwan (R.O.C.) 10617
`shukaihsieh@ntu.edu.tw`

Abstract

In the Generative Lexicon Theory (GLT), co-composition is one of the generative devices proposed to explain the cases of verbal polysemous behavior where more than one function application is allowed. The English baking verbs were used as one of the examples to illustrate how their complements co-specify the verb with *qualia unification*. In this paper, we begin by exploring the polysemy of Chinese baking verb, where the first two senses in Chinese Wordnet (CWN) are assumed. Features including linguistic cues and common sense knowledge are involved in the experiment with Weibo corpus and computed with SVM for closer investigation. From the analysis, it is found that though there are various cases found in senses of *change of state* and *creation*, a coarse but systematic approach combined with certain features in disambiguating CWN senses could be arranged. In addition, we further observe that the usage of various instruments cases and classifiers would be harnessed by underlying background knowledge to help select an appropriate sense based on the context.

Keywords: The generative lexicon, co-composition, baking verbs

1 Introduction

In Generative Lexicon Theory (GLT), the co-composition theory in discussing the logical polysemy of verbs illustrates that in some verbal meaning alternations, arguments of verbs would shift the meaning of verb in the compositional interpretation. This poses difficulties for word sense disambiguation (WSD) task in contextualizing the underlying sense, i.e., putting semantic weights on the non-functor elements, to give rise to a derivative sense.

Firstly, we start with exploring the representative example of baking verb "bake" used in GLT regarding co-composition, to see whether two assumed senses, *change of state* and *creation*, can be derived through the proposed generative mechanisms in composition with its argument in the case of Chinese baking verb *kao* 烤 'bake' with the Chinese examples *kao malingshu* 烤馬鈴薯 'bake a potato' and *kao dangao* 烤蛋糕 'bake a cake'.

We choose WordNet to depict the contrast. In regard to the differences in senses *change of state* and *creation* of the English verb "bake", the definition of the verb in WordNet [1], carry diverse glosses as well in the examples "bake a potato" and "bake a cake"; however, considering the verb *kao* 烤 'bake' in Chinese WordNet (CWN) [2], both the Chinese examples *kao malingshu* 'bake a potato' and *kao dangao* 'bake a cake' could be included into the first CWN gloss "use heat to cook and make the food edible" (CWN_sense_1). Whereas, it is discovered that the example *kao malingshu* 'bake a potato' could also be applied to the secondary CWN gloss "use heat to heat the object" (CWN_sense_2). That is to say, in Chinese, although *change of state* sense would be assigned to *kao malingshu* 'bake a potato' and *creation* sense would be attached to *kao dangao* 'bake a cake' as in English, both examples would be primary grouped into CWN_sense_1; but there are situations that *kao malingshu* 'bake a potato' also occur with CWN_sense_2, based on the context. Therefore, in the above Chinese cases, it seems to be clear that examples with *creation* sense would only be assigned to CWN_sense_1; while the con-

---

ditions for examples with *change of state* sense, to distinguish interpretation differences between CWN_sense_1 and CWN_sense_2, need to be further investigated.

Nonetheless, there are situations that examples with *creation* sense would be assigned to CWN_sense_2 as well. For instance, *kao tusi* 烤土司 'toast a loaf of bread / toast a slice of toasted bread' would not only be assigned to CWN_sense_1, but also CWN_sense_2, based on some occasions. Additionally, a sense shifting would be prompted as well, from *creation* sense to *change of state* sense. Moreover, it is investigated that cases such as *kao dofu* 烤豆腐 'grill tofu' though along with *change of state* sense, yet possesses some features from *creation* sense, and could merely be specified with CWN_sense_2.

Therefore, this paper aims to search out a coarse but systematic approach with linguistic cues, with the help of applying the Support Vector Machine (SVM) computational technique by taking Leidon Weibo Corpus (van Esch, 2012) [3], to help identify and analyze what the sets of conditions for *change of state* and *creation* senses are that would lead to different mappings between CWN_sense_1 and CWN_sense_2, by investigating the Chinese baking verb *kao* 'bake'.

2   Co-composition in GLT

Qualia structure (Pustejovsky, 1995), adapted from the modes of explanation by Aristotle, depicts that there are four main essential factors (*constitutive*, *formal*, *telic*, and *agentive*) to drive and capture the interpretation of an object as well as a relation (Moravcsik, 1975). Although many models of semantics agree that words have simple denotations, but there are various perspectives in the methods of lexical composition. Some formal models argue that the composition approaches are truth-value denotation and computation within logical inferences; while in the perspective of GLT, it is the semantic transformations (including type coercion, selective binding, and co-composition) of words' denotations that shift from one to another to form new meanings. Therefore, in GLT, the use of qualia structure could be applied to better specifying a word's meaning.

As mentioned in Pustejovsky (1995), among the four interpretive levels of qualia structure, the *agentive* quale of the lexical item is encoded with the knowledge of what an object may identify or refer to and be able to explain an artifact comes into being. Therefore, it would be an important manner if something is created in order to distinguish natural kinds (e.g. potatoes, carrots and so on) from artifacts (e.g. cookies, cakes, bread).

In addition, the *agentive* role of a lexical item would be represented as an event predicate while the lexical item is a noun. For example, "potato" and "cake" could all be event predicates in "bake a potato" and "bake a cake"; however, the verb "bake" is polysemous with two meanings: a sense of *change of state* and a sense of *creation*, as stated in Atkins et al. (1988). Since this kind of logical polysemy occurs in many cases, a relation of co-composition is introduced by Pustejovsky (1995) (originally named as co-specification (Pustejovsky, 1991)) to capture the words' meanings.

Under the notion of co-composition, the verb "bake" itself is not polysemous but the complement that follows derives other meanings can be re-examined, not only through the *agentive* quale, but also *constitutive* role. From the example (51) provided by Pustejovsky (1995), it is further discovered that though a complement makes reference to an *agentive* quale, the *constitutive* quale plays an important role to the baking act. That is, if the material in a *constitutive* quale of a complement is an individual as a default argument, the derived sense from *agentive* role would be *change of state*. On the other hand, when the material in a *constitutive* quale of a complement is a mass of individual components, the selected sense from *agentive* role would turn out to be *creation*.

Therefore, the verb "bake" originally has one event type but with two argument types in the lexical structure, it is the complement that chooses one of the two arguments to govern. When in the case of "bake a potato", the *agentive* role of "potato" is simply a natural kind and an individual material, the process only involves state changes with event type makes no shifting, and thus the sense *change of state* is assigned; whereas in "bake a cake", "cake" is an artifact created from a mass of components, its event type would shift from one to the other, thus obtain the sense of *creation*. The kind of event type shifting in a complement, is what that

makes the verb "bake" to be polysemous, not the verb itself.

The co-composition operation on VP proposed by Pustejovsky (1995) includes the following process:

1. The governing verb would apply to its complement;

2. The complement would then co-specify the verb;

3. A new sense of the verb would be derived resulting from an operation called *qualia unification*, where the agentive roles of both the verb and its complement match with each other; and the *formal* quale of the complement is also the *formal* role of the entire VP.

Since the process of co-composition will arouse new senses to the governing verb based on its complement, it is also worth noting that the thematic roles played by the complement of a verb needs to be taken into consideration. As mentioned by many researchers (Tanenhaus et al., 1989; Jackendoff, 1987; Gentner, 1981), thematic role knowledge is part of a verb's meaning and can be construed as a claim that the concept of a verb is its relation to the entities participated in the event. Though numerous thematic taxonomies have been proposed by linguists, six thematic roles are typically involved: agent, patient, theme, goal, instrument and location (Cook, 1979; Fillmore, 1968).

In this paper, we are interested in investigating the verb *kao* 'bake' with its complements in Chinese examples under co-composition theory, and focus on exploring one of the thematic roles "instrument" within context. This study thus aims to further seek empirically for what the linguistic features are in deciding or shifting the CWN senses of *kao* 'bake' under the notion of *change of state* and *creation* senses. A corpus-based machine learning approach is taken for the analysis, which is introduced in the following section.

3 Data Analysis using Weibo Corpus

3.1 Data Collection

Since Weibo is the most prevalent Chinese social communication and microblogging platform, recent studies in corpus data analysis have taken Weibo data as corpus, in order to further and better investigate the up-to-date language usage. By

taking the Weibo corpus into study, we can not only freely accessed large amount of timely data without expensive computing, but also discover the linguistic cues that best display current language usage. Therefore, in this paper, with the open-sourced weibo corpus, Leiden Weibo Corpus (van Esch, 2012), freely accessed online, the posts containing *kao* 'bake' could all be easily retrieved using R programming language. At present, due to the efficiency in data processing, convenience in applying statistical models and powerfulness in plotting, an amount of 9688 parsed posts involving the verb *kao* 'bake' have been successfully extracted for the preparation of following data analysis.

3.2 Complements, Linguistic Features and Common Sense Knowledge

By observing the extracted *kao* 'bake' posts, 53 nouns that could be taken as complements of the verb are randomly chosen, and manually tagged with one of the two senses based on its complement role to the verb *kao* 'bake'(e.g. 41 nouns are tagged as *change of state* sense and the other 12 are tagged as *creation* sense), as target data for running SVM approach. Since complements may trigger VPs to select a *change of state* or *creation* sense, there might be some certain embedded and underlying information (including linguistic cues and common sense knowledge) beneath a complement, which causes the complement to select one of the two senses for an VP. Therefore, by applying SVM approach with related implicit information of complements, we may roughly further investigate what the information are influencing the decision of senses between *change of state* and *creation*. In addition, via the analysis of SVM results, a more detailed exploration can be carried out from the observed essential implicit information, in determining CWN_sense_1 and CWN_sense_2 under *change of state* and *creation* senses.

Hence, a data frame targeted on the complements with features such as relevant linguistic cues and common sense knowledge, to learn whether these features would help deciding *change of state* and *creation* senses, is shown in Figure 1. We tend to add in as many relative linguistic cues and shared knowledge as possible, and by using the characteristics of SVM to help quickly derive a set of effective features from a pool of various infor-

Figure 1: A data frame with complements and features prepared for applying SVM approach

mation for further observation.

In the data frame, since the collocation of a word has been taken as an approach in computational linguistics for presenting its relationships with a word, the collocations are also involved for SVM. The span of the collocation is set to three before the position of the nouns for automatically extracting and computing, and the first collocation method in Gries (2009) is applied. Since it is observed in GLT that the sense of a verb would be influenced by the followed noun, despite the verb *kao* 'bake', it would be interesting to see whether there are shared or common interactions between the complements and other verbs that follows. In addition, studies of classifiers have shown to be an important feature in representing a noun and have been applied to various classifiers to help make divisions. Therefore, in regard to linguistic features adapted in this paper, only the collocations of each noun with the highest frequency counts in verbs and classifiers, will be selected. The 16 selected linguistic features (including 9 verbs and 7 classifiers) are computed and each given a scaled[4] Point-wise Mutual Information (PMI) value to each noun. Equation is shown in (1).

$$PMI = \frac{P(X,Y)}{P(X) * P(Y)} \quad (1)$$

Unlike the 16 linguistic cues that could be retrieved easily from the data, the underlying common sense knowledge is hard to be revealed. Accordingly, 5 common sense features are manually analyzed and tagged, which include animacy (animate or inanimate), artifact (an artifact or natural objects), culinary (needed to be cooked before eating or not), sweets (could be generally categorized as sweets or not), and constitutive (whether its default argument is an individual or a mass).

3.3 Data Training and Testing

In order to see the interactions between the 21 features and the 53 nouns, the SVM approach is introduced to investigate whether the features mentioned above could possibly provide sufficient information of a complement to select *change of state* or *creation* sense for the verb, and furthermore, make a further research on finding essential features from the potential information in helping disambiguate CWN_sense_1 and CWN_sense_2. Therefore, the data in the data frame is randomly divided into two groups, which 70% of the data is used for training a model and the rest of the 30% is for testing. Furthermore, the results of SVM is presented by involving F-score, see equation (2), to present a weighted average of the precision and

---

[4]The numbers shown in Figure 1, are all scaled by applying collocation frequencies to Z-score, in order to get the data weighted for a better investigation. Therefore, the scaled numbers would have positive and negative values.

recall, and the score ranges from 0 (the worst) to 1 (the best).

$$F = 2 * \frac{precision * recall}{precision + recall} \quad (2)$$

4 Analysis

From SVM approach, the F-score presents 0.67 value for the model. Although the F-score only shows 67% chance to correctly make the complement choose the right *change of state* and *creation* senses for the verb *kao* 'bake', some inconsistency could be found within the 5 manually tagged common sense knowledge features and could be considered for the further discussion in dividing CWN_sense_1 and CWN_sense_2.

- Animacy - For complements that are inanimate are all tagged with *creation* sense; however, there are some that would be grouped as *change of state* sense.

- Sweets - For complements that are not sweets are all tagged with *change of state* sense; however, some would be assigned with *creation* sense.

- Culinary - For complements that do not needed to be cooked before eating, are tagged with *creation* sense; however, some would be fixed with the *change of state* sense.

- Artifact and Constitutive - It is found that the tags between artifact and constitutive are consistent. This might lead to the reason that if an item is an artifact instead of natural objects, a lot of materials would included for an artificial process. Therefore, for complements that are assigned as artifacts, are also tagged as mass; and vice versa. In addition, for complements that are tagged as artifact and mass, are all tagged with *creation* sense; however, some would be categorized as carrying *change of state* sense.

As observed from the above features, there are some complements containing the characteristics of being a *creation* sense, but are assigned with the sense of *change of state*. Though features change along with the senses, a typical combination of deriving a *creation* sense based on the complement could still be found, which including features such as inanimate, a kind of sweets, not culinary, an artifact and coming from a mass of materials.

As mentioned in GLT, the *constitutive* quale of whether a material is individual or a mass within a noun, would help identify a sense for a VP. Thus, as the examples described in Pustejovsky (1995), when the *constitutive* role is an individual, the sense of *change of state* is chosen; and when the *constitutive* role is a mass, the sense of *creation* would be assigned. However, in Chinese, it is found that there are examples that would be specified as *change of state* sense when the *constitutive* role is a mass. More examples and illustrations will be presented in the following section, and a brief process in identifying CWN_sense_1 and CWN_sense_2 under *change of state* and *creation* senses, are presented in Figure 2.

4.1 *Change of state* sense

4.1.1 *Constitutive* role: individual

By applying the *constitutive* quale, Chinese examples with constitutive quale identified as an individual, would be mapped to CWN_sense_1. These examples observed from the corpus, with the state changing from raw to cooked, could be roughly categorized as three groups: meat (e.g. *kao niurou* 烤牛肉 'roast beef' and *kao yangrou* 烤羊肉 'grill mutton'), seafood (e.g. *kao yu* 烤魚 'grill fish' and *kao longxia* 烤龍蝦 'grill lobsters'), and vegetables (e.g. *kao malingshu* 烤馬鈴薯 'bake a potato' and *kao xianggu* 烤香菇 'grill mushrooms').

4.1.2 *Constitutive* role: mass

For those examples with *change of state* sense but possess the *constitutive* role as a mass, which is one of the features to be specified as *creation* sense, observed from the corpus are *kao daofu* 烤豆腐 'grill tofu', *kao xiangchang* 烤香腸 'grill sausages', *kao mianjin* 烤麵筋 'grill gluten', *kao regou* 烤熱狗 'grill hotdogs', *kao jiu* 烤酒 'heat liquor', *kao chunjuan* 烤春捲 'grill spring rolls', *kao boazi* 烤包子 'grill steamed buns', and *kao sanmingzhi* 烤三明治 'grill sandwiches'.

Mostly these examples would only be led to CWN_sense_2, with the state changing from cold/cool to heated; however, this is not the case when considering *kao xiangchang* 'grill sausages' and *kao regou* 'grill hotdogs', which could be assigned to CWN_sense_1 as well.

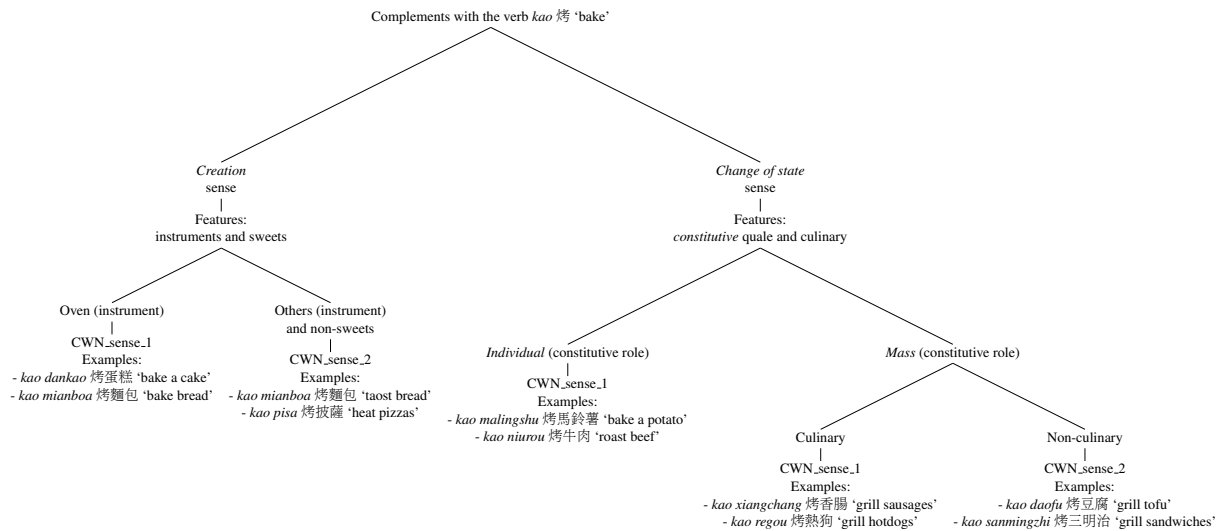Taking *kao xiangchang* 'grill sausages' for instance, as presented in example (3):

Complements with the verb *kao* 烤 'bake'

Creation
sense
Features:
instruments and sweets

Change of state
sense
Features:
constitutive quale and culinary

Oven (instrument)
CWN_sense_1
Examples:
- *kao dankao* 烤蛋糕 'bake a cake'
- *kao mianboa* 烤麵包 'bake bread'

Others (instrument)
and non-sweets
CWN_sense_2
Examples:
- *kao mianboa* 烤麵包 'taost bread'
- *kao pisa* 烤披薩 'heat pizzas'

Individual (constitutive role)
CWN_sense_1
Examples:
- *kao malingshu* 烤馬鈴薯 'bake a potato'
- *kao niurou* 烤牛肉 'roast beef'

Mass (constitutive role)

Culinary
CWN_sense_1
Examples:
- *kao xiangchang* 烤香腸 'grill sausages'
- *kao regou* 烤熱狗 'grill hotdogs'

Non-culinary
CWN_sense_2
Examples:
- *kao daofu* 烤豆腐 'grill tofu'
- *kao sanmingzhi* 烤三明治 'grill sandwiches'

Figure 2: The process of identifying CWN_sense_1 and CWN_sense_2 under *change of state* and *creation* senses

(3) 回家　　　　　高速路　上　　　的
huijia　　　gaosulu shang de
on the way home freeway SHANG DE
休息站　　開始　賣　烤　香腸
xiuxizhan kaishi mai kao xiangchang
rest area start sell grill sausage
和　烤　肉丸　　了
he kao rouwan le
and grill meat ball LE
'On the way home, the rest area beside the freeway, starts to sell grilled sausages and meat balls.'

As presented in example (3), though the *constitutive* role of sausages would be specified as a mass by containing a lot of ingredients, it is CWN_sense_1 that would be assigned to rather than CWN_sense_2. Such cases could be re-analyzed and distinguished by the manually tagged feature: culinary. For cases that are tagged as culinary, which illustrates "need to be cooked before eating", would then be grouped as CWN_sense_1; whereas, those that are tagged as non-culinary, expressing "edible without being cooked", would be specified as CWN_sense_2. Therefore, since *kao xiangchang* 'grill sausages' in example (3) is identified as culinary, it would be directed to CWN_sense_1.

4.2 *Creation* sense

4.2.1 Using instrument: oven

Considering the examples carrying *creation* sense: *kao dangao* 烤蛋糕 'bake a cake', *kao bing-gan* 烤餅乾 'bake cookies', *kao gaobing* 烤糕餅 'bake pastries', *kao tiantianquan* 烤甜甜圈 'bake donuts', *kao danjuan* 烤蛋捲 'bake egg rolls', *kao subing* 烤酥餅 'bake shortcakes', *kao buding* 烤布丁 'bake puddings', *kao mianboa* 烤麵包 'bake bread', *kao shaobing* 烤燒餅 'bake sesame seed cakes', *kao tusi* 烤土司 'bake a loaf of bread', *kao xiang* 烤饢 'bake a kind of traditional bread from north China', and *kao pisa* 烤披薩 'pizzas', it is investigated that these cases specified as CWN_sense_1 all share the same feature, using oven as instrument.

One of the example *kao dangao* 'bake a cake' is used for the following illustration.

(4) 沒　注意　看　　　烤箱,　蛋糕
mei zhuyi kan kaoxiang dangao
not notice watching oven cake
烤　　過了頭
kao guoletou
bake overtime
'Not noticing the oven, the cake is over-baked.'

Therefore, as presented in example (4), it is the use of instrument *kaoxiang* 烤箱 'oven' that frequently follows when cases that are tagged as *creation* sense.

In addition, the use of oven in a *creation* sense among the examples with the verb *kao* 'bake' becomes a common sense in shared knowledge. For example in *kao dangao* 'bake a cake', it may be the bakery bakers, not anyone else, that would frequently use 'oven' to bake a cake. Hence, the

stereotype of using an oven for baking a cake is then implanted into the mind as background knowledge. That is to say, even without the instrument 'oven' occurred in the context, the act of baking a cake already possesses the default information of instrument 'oven', and thus would still be assumed as carrying the meaning CWN_sense_1, as shown in example (5).

(5) 我們　　的　　家庭　　生活*!*　　周末
     women　de　jiating　shenghuo　zhoumo
     our　　DE　family　life　　　weekend
     烤　　蛋糕　　嘍*!*
     kao　dangao　lou
     bake　cake　LOU
     'Our family life! Bake a cake during weekend!'

However, there are still some examples that take not only the oven as the only instrument, but others such as toasters, grills and so on, would be assigned to CWN_sense_2 with *change of state* sense in some occasions. The discussion about these would leave to section 4.2.2 for more details.

For those that only use oven as instrument are typical examples with *creation* sense, which would not also possess the *change of state* sense depending on the context. Examples that meet with the requirements are: *kao dangao* 'bake a cake', *kao binggan* 'bake cookies', *kao gaobing* 'bake pastries', *kao tiantianquan* 'bake donuts', *kao danjuan* 'egg rolls', *kao subing* 'bake short-cakes', and *kao buding* 'bake puddings'. Furthermore, it could be inspected that these cases not only share the feature of merely using oven as instrument, but also are all consistently manually tagged as sweets, which is one of the 5 manually tagged common sense features. Therefore, if an example with *creation* sense is investigated to be a kind of sweets and only uses oven as instrument for baking, it could then be directly grouped to CWN_sense_1.

4.2.2　Using instrument: others

For examples that are tagged with *creation* sense and are not sweets, might also possess *change of state* sense with CWN_sense_2 depending on the context, such words are *kao mianboa* 'toast bread', *kao shaobing* 'heat sesame seed cake', *kao tusi* 'toast a slice of toasted bread', *kao xiang* 'heat a kind of traditional bread from north China', and *kao pisa* 'heat pizzas'.

The following takes the example *kao mianboa* 'toast bread' for illustration.

(6) 眼看　　烤箱　　　裡　　的
     yankan　kaoxiang　li　　de
     see　　oven　　　inside　DE
     羊角麵包
     yangjiaomianbao
     croissant
     'See the croissant inside the oven'

(7) 轉載　　　我　用　　東菱　　　麵包機
     zhuanzai　wo　yong　dongling　mianbaoji
     forward　I　　use　　Donlim　　toaster
     烤　　麵包　　的　配方　　　　與
     kao　mianbao　de　peifang　　　yu
     toast　bread　DE　cooking recipe　and
     步驟
     buzou
     step
     'Forward the cooking recipe and steps that I use for toasting bread on the Donlim toaster.'

As presented in example (6) and (7), the instrument for bread could either be an oven or a toaster. The sense in example (6) stays as what it is originally tagged, the *creation* sense with the usage of instrument 'oven'; whereas in example (7), when the instrument is other than an oven, such as a toaster, *change of state* sense would then be selected. This could lead to the reason that the instrument, toaster, is mainly used for heating the bread, rather than baking or creating the bread. Thus, the process of applying the instrument toaster to perform the act *kao mianboa* 'toast bread', would simply be *change of state* sense, along with state changes from cool/cold to heated.

Besides, by considering whether examples are using an oven as instrument or not, classifiers may provide some contributions in helping identifying the occasions that cases with *creation* sense would become *change of state* sense.

(8) 早餐　　　吃　　了　兩　　片　　烤
     zaocan　　chi　le　liang　pian　kao
     breakfast　eat　LE　two　　piece　toast
     土司
     tusi
     bread
     '(Someone) eats two pieces of toasted bread for breakfast.'

(9) 他 烤 了 一 條 牛奶土司
　　ta　kao　le　yi　tiao　niunaitusi
　　he　bake　LE　one　loaf　milk bread
　　'He bakes a loaf of milk bread.'

As shown in example (8) and (9), different classifiers also implicate the using of certain instruments that may implicitly select a *change of state* sense for example (8) and a *creation* sense for example (9). Though the instruments are not revealed in the two examples, the classifier *pian* 片 'piece' in example (8), indicated the implicit instrument 'toaster' which would usually be used for toasting slices of bread or toast. Hence, due to the underlying usage of a toaster instead of an oven, the *change of state* sense could be affirmed. Considering the example (9), using the classifier *tiao* 條 'loaf' suggests that it is often the instrument oven that would bake a loaf of bread. On that account, the *creation* sense could be verified.

5 Conclusion

Under the point of co-composition, in order to observe an approach with linguistic cues that influence a complement to select a *change of state* or *create* sense for the Chinese baking verb *kao* 'bake', the investigation and analysis are carried out by using Leiden Weibo Corpus along with the application of SVM technique.

From the analysis, it is figured out that the sense of a complement with the verb *kao* 'bake' might be influenced by two of the five manually tagged features: sweets and culinary, usage of instruments and *constitutive* quale.

In Chinese examples, when a complement follows the verb *kao* 'bake', conditions for assigning *change of state* or *creation* sense to a CWN_sense_1 or CWN_sense_2 are as below: [1] *change of state* sense: if the *constitutive* role of a complement is an individual, or a mass but tagged as culinary, a CWN_sense_1 would be assigned; while if its *constitutive* role is a mass and tagged as not culinary, a CWN_sense_2 would be chosen. [2] *creation* sense: if the instrument oven for performing the act of *kao* 'bake', then a CWN_sense_1 would be selected; however, if using the instrument other than oven and not being tagged as sweets, a CWN_sense_2 would be specified.

Therefore, in *change of state* sense, by combing the *constitutive* role and culinary feature, the CWN_sense_1 and CWN_sense_2 could be identified; whereas, in *creation* sense, with the help of instrument usage and sweets feature, the situations by assigning CWN_sense_1 or CWN_sense_2 could be affirmed. Moreover, if the instrument is omitted, the classifiers could further help decide which sense to be assigned to under the situation.

As could be observed from the results, it seems that most of the CWN_sense_1 could be shifted to CWN_sense_2 according to the context. This is due to the reason that the *change of state* sense in Chinese examples would have two meanings: [1] state changes from raw to cooked, which is CWN_sense_1; [2] state changes from cool/cold to heated, which is CWN_sense_2. Thus, since most state changes in Chinese need to be firstly altered through the process in CWN_sense_1, the situations that cooked food get cool/cold and would like to get heated, have been discovered in context.

Consequently, in this paper, inspired by the analysis of discussing English baking verb "bake" under co-composition theory, we take this as a starting point for a preliminary study in a specific sub-task of Chinese Word Sense Disambiguation (WSD). Future works include extending the model to handle other underspecified phenomena, e.g. creation verbs and performance verbs, where information from complements and other non-functor elements co-compose to give rise to derived sense.

References

B. T. Atkins, J. Kegl, and B. Levin. 1988. Anatomy of a verb entry: From Linguistic Theory to Lexicographic Practice. *International Journal of Lexicography*, 1:84–126.

W. A. Cook. 1979. *Case grammar: The development of the matrix model (1970V1978)*. Washington: Georgetown University Press.

C. Fillmore. 1968. *The case for case*. New York: Holt, Rinehart and Winston.

D. Gentner. 1981. Some interesting differences between nouns and verbs. *Cognition and Brain Theory,*, 4:161–178.

Stefan Th. Gries. 2009. *Quantitative Corpus Linguistic with R: A Practical Introduction*. London & Newyourk: Routledge, Taylor & Francis Group.

R. Jackendoff. 1987. The status of thematic relations in linguistic theory. *Linguistic Inquiry*, 18:369–411.

J. Michael Moravcsik. 1975. Aitia as generative factor in aristotle's philosophy. *Dialogue*, 14:622–636.

James Pustejovsky. 1991. The generative lexicon. *Computational Linguistics*, 17(4):409–441.

James Pustejovsky. 1995. *The Generative Lexicon*. The MIT Press.

M. K. Tanenhaus, G. N. Carlson, and J. T. Trueswell. 1989. The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes*, 4:211–234.

Daan van Esch. 2012. Leidon weibo corpus. http://lwc.daanvanesch.nl/index.php.