ICIC Express Letters
Part B: Applications
Volume **7**, Number **7**, July **2016**

ICIC International ©2016 ISSN 2185-2766
pp. **1439–1444**

LEXICAL-SEMANTIC CONSTRAINTS ON THE SEMANTICS-TO-SYNTAX PROJECTING: A CASE STUDY OF LOCATION

Shiyong Kang[1,2], Minghai Zhou[1,2] and Linlin Shi[1,2]

[1]Key Laboratory of Language Resource Development and Application of Shandong Province
[2]School of Chinese Language and Literature
Ludong University
No. 186, Middle Hongqi Road, Zhifu District, Yantai 264025, P. R. China
kangsy64@163.com

Received December 2015; accepted March 2016

Abstract. *The lexicon-to-syntax projecting is semantically constrained. This article exemplifies the semantic role Location, focusing on its morphological marker, syntactic components it can be projected as, the influence of its subcategories on the projecting, and its semantic characteristics, based on a large-scale annotated corpus, and hence gives a preliminary illustration on how the semantic categories constrain the lexicon-to-syntax projecting.*
Keywords: Semantic role, Syntactic component, Semantic category

1. **Introduction.** A word serves as a certain semantic role once it appears in a sentence [1]. The projecting of a semantic role to a corresponding syntactic component is constrained by its semantic meaning [2-7]. Financially supported by the National Grant of Social Science, a study on the semantic constraint on the semantic-to-syntactic projecting, a large-scale corpus had been annotated and some rules about how a semantic role is projected as a syntactic element had been drawn, which as our preliminary work facilitates this study.

As for the semantic category, the classifying and labeling systems of *Tongyici Cilin* [8] are adopted in this study. This paper summarizes the constraints on the projection of the semantic role *Location* onto syntactic elements based on this annotated corpus.

Scholars differ greatly in naming and classifying semantic roles, and even the same scholar may use different names or have different classifications for semantic roles in different periods or in different research papers or books. So far *Location* has been mentioned by some scholars and institutions, including Shao [9], Lin and Lu [10], Lu [11], Sinica Treebank [12], and HowNet [13].

In our framework, *Location* is determined as the place where events take place, situation and path of event. For instance, in the sentence "*huamei* (throstle) [D *zai* (at) *shulin* (woods) *bian* (side)] *wanzhuan de* (beautifully) [P *changge* (is singing)]", "*zai* (at) *shulin* (woods) *bian* (side)" is the location role semantically and an adverbial modifier syntactically of the predicate "*changge* (is singing)". We annotated the syntactic components and corresponding predicate verbs for the location roles in each sentence of the texts of the Chinese teaching material for elementary and secondary school published by People Education Press, and then performed detailed analysis based on this corpus.

The rest of this paper is organized as follows. Section 2 describes the morphological markers of *Location*. Sections 3 and 4 analyze the syntactic components of *Location* and the lexical-semantic constraints on the projection of *Location*, respectively. The characteristics of *Location* are discussed in Section 5. Finally, a brief conclusion is given in Section 6.

2. **Morphological Marker of *Location*.** [14] points out that the semantic role *Location* is characterized by a certain morphological marker, including location marker and naming marker. The former are the location words, such as *shang* (on, upper), *xia* (under, below), *zuo* (left), *you* (right), *li* (inside), *wai* (outside), *nei* (in, among), *zhouwei* (around, about), and *zhong* (amidst). Location words are divided into two classes, prepositional localizer and postpositional localizer. The former marks the connection, and the latter the attributes.

Table 1. Postpositional localizer

Localizer	Subject	Complement	Adverbial	Object	Total
bian (near)	0	0	0	5	5
wai (outside)	14	0	6	6	26
zhong (amidst)	35	51	179	33	298
shang (upper, on)	155	77	240	73	545
zhijian (in between)	7	8	4	3	22
jian (at the tip of)	2	0	0	0	2
xia (below, under)	38	49	51	15	153
dixia (under)	4	1	0	4	9
qian (in the front of)	33	58	16	11	118
hou (behind)	29	49	3	3	84
pang (by the side of)	13	21	13	2	49
dangzhong (in the middle of)	1	0	0	0	1
fujin (nearby)	1	0	0	0	1
nei (within, inside)	4	0	6	0	10
jian (between, among)	11	0	0	0	11
chu (place)	17	0	0	0	17
zhizhong (in the midst of)	1	0	0	0	1
li (inside)	32	29	274	155	490

Table 2. Prepositional localizer

Localizer	Subject	Complement	Adverbial	Object	Pivotal Object	Total
zai (at)	16	670	843	192	2	1723
cong (from)	1	2	265	0	3	271
xiang (towards)	1	0	44	0	0	45
dao (reach)	1	78	18	424	0	521
hui (back)	0	1	0	58	0	59
yu (at)	1	43	2	12	0	103

3. **Localizers and the Corresponding Syntactic Components.** *Location* role is allowed to serve as five types of syntactic components, i.e., adverbial modifier, object, complement, subject and pivotal object, as shown in Table 3. The statistical results suggest that *Location* can serve as object, complement, subject and concurrent phrase, besides as the adverbial modifier.

3.1. **Regular syntactic component.**

a). Projected as adverbial modifier (D), 1512 in all, proportion: 33.56%. For example:
[D *Wo* (I) *zou* (walk) *bu* (not) *guoqu* (pass) *de* (de) *difang* (place)] *ni* (you) *jiu* (may) [P *beizhe* (carry)] *wo* (me).

b). Projected as object (O), 1268 in all, proportion: 28.14%. For example:
Ta (He) *tingxialai* (stopped), (,) [P *huangu* (look about)] [O *sizhou* (all round)].

Table 3. Frequency distribution of five syntactic components

Syntactic component	Frequency	Proportion
Subject	703	15.6%
Complement	937	20.79%
Adverbial	1512	33.56%
Object	1268	28.14%
Pivotal object	1	0.02%

3.2. **Irregular syntactic component.**

a). Projected as subject (S), 703 in all, proportion: 15.6%. For example:

[P *Feiteng* (boiling) *le* (le)], (,) [S *cunzi li* (the village)].

b). Projected as complement (C), 937 in all, proportion: 20.79%. For example:

Wode (my) *qizi* (wife) *he* (and) *erzi* (son) [P *zou* (walk)] [C *zai* (in) *houmian* (the back)].

c). Projected as pivotal object (J), 1 only, proportion: 0.02%. For example:

Tade (his) *caichan* (property) *de* (of) *banshu* (half) *yingdang* (should) [P *gui* (belong to)] [J *shouhai de* (damaged) *yifang* (party)] [P *suoyou* (own)].

4. Influence of Semantic Category in Projecting.

4.1. **Semantic categories of *Location*.** *Location* can be projected as more than one syntactic component, but the projecting competence varies cross semantic categories. As illustrated in Table 4, these four semantic categories are in the same descending scale to be projected as the syntactic components as follows:

Serving as adverbial modifier: C > B > D > A

Serving as complement: C > B > D > A

Serving as object: C > B > D > A

Table 4. Semantic distribution of different syntactic components at category level

Semantic Category	Subject	Complement	Adverbial	Object
A (person)	7	2	6	4
B (object)	221	180	326	236
C (time space)	313	691	826	323
D (abstract things)	109	94	120	88
Total	703	937	1512	1268

As seen from above, the category C shows the most powerful competence to be projected as any syntactic element. The reason might be that there are many reference words like *zher* (here), *nar* (there), z*heli* (here), and *nali* (there). The less powerful is the category B. For example:

[S *libian* (inside)] [P *mei* (do not exist)] *ren* (person)

Libian (inside) in the above example belongs to C respectively.

To refine the statistical study, we went further to discuss the semantic subcategories, as shown in Table 5.

As seen above, these subcategories serving as different syntactic components are ranked as follows:

Subject: Cb (Space) (313) > Bk (The whole body) (126) > Di (Social politics and law) (109) > Bn (Buildings) (79) > Bc (Object parts) (56) > Bm (Material) (33) > Bf (Weather) (29) > Be (Geomorphology) (27) > Bd (Celestial body) (13) > Aa (General term) (7) > Bp (Appliance) (6) > Bg (Natural objects) (4)

Complement: Cb (Space) (691) > Bk (The whole body) (109) > Di (Social politics and law) (94) > Be (Geomorphology) (88) > Bn (Buildings) (85) > Bc (Object parts) (66) > Bm (Material) (43) > Bf (Weather) (37) > Bd (Celestial body) (10) > Bo (8) > Bg (Natural objects) (7) > Bp (appliance) (2)

Adverbial: Cb (Space) (936) > Di (Social politics and law) (120) > Bk (The whole body) (98) > Bn (Buildings) (77) > Bc (Object parts) (60) > Be (Geomorphology) (32) > Bm (Material) (27) > Bf (Weather) (26) > Bg (Natural objects) (15) > Bo (Machine and tools) (11) > Bp (appliance) (11) > Bd (Celestial body) (9)

Object: Cb (Space) (323) > Bk (The whole body) (97) > Di (Social politics and law) (88) > Bc (Object parts) (79) > Bn (Buildings) (65) > Bm (Material) (54) > Bg (Natural objects) (41) > Bd (Celestial body) (28) > Bf (Weather) (19) > Bp (appliance) (9) > Be (Geomorphology) (8) > Bo (Machine and tools) (5)

TABLE 5. Semantic distribution of different syntactic components at subcategory level

Subcategory	Subject	Complement	Adverbial	Object	Subcategory	Subject	Complement	Adverbial	Object
Aa	7	2	6	4	Ab	/	/	/	/
Ba	1	/	3	/	Bb	/	2	/	/
Bc	56	66	60	79	Bd	13	10	9	28
Be	27	88	32	8	Bf	29	37	26	19
Bg	4	7	15	41	Bh	/	/	/	/
Bk	126	109	98	97	Bl	/	/	/	/
Bm	33	43	27	54	Bn	79	85	77	65
Bo	4	8	11	5	Bp	6	2	11	9
Ca	/	/	/	/	Cb	313	691	936	323
Di	109	94	120	88	Dj	/	/	/	/

From above it is clearly observed that Cb is prior to all the other subcategories to be projected as any syntactic component, which is exemplified as follows:

Adverbial: *Women de* (our) *yiqie* (any) *huai* (bad) *xiaoxi* (message) *dou* (all) *shi* (is) [D *cong* (from) *nali* (which)] [P *chuan* (spread) *chulai* (out)] *de* (de).

4.2. **Semantic categories of predicate selecting *Location*.** The semantic categories of the predicates which are selected by *Location* are illustrated as in Table 6, from which it is concluded that these semantic categories of verbs are ranked as:

F (action) > H (activity) > J (connection) > G (psychology) > K (assistant phrase) > I (phenomena and states) > E (feature)

TABLE 6. Semantic distribution of predicate verbs

Category	Subject	Complement	Adverbial	Object	Predicate	Total
E (feature)	0	0	1	0	0	1
F (action)	425	339	659	511	3	1937
G (psychology)	5	15	32	42	0	94
H (activity)	309	510	559	498	2	1878
I (phenomenon and state)	0	0	2	3	0	5
J (connection)	39	43	55	65	0	202
K (assistant phrase)	4	2	38	32	0	76
L (honorific)	0	0	0	0	0	0
Total	782	909	1346	1151	5	4193

5. **Characteristics of *Location*.** *Location* usually serves as adverbial modifier, but it is also allowed to occur in other positions. The point is that in any position, Cb is the most optimal subcategory to be projected, which confirms our hypothesis that *Location* is assumed by location word. *Location* falls into abstract location and concrete location. The former is invisible and untouchable, while the latter features the opposite, which can be divided into three classes.

1) Concrete place name

Wo (I) [D *cong* (from) *Beijing* (Beijing)] [P *dao* (went to)] [O *Xuzhou* (Xuzhou)].

2) Object

Zheyang (Thus), (,) *women* (we) [D *zai* (under) *yangguang* (sunshine) *xia* (below)], (,) [D *xiangzhe* (to) *na* (that) *caihua* (cauliflower), (,) *sangshu* (mulberry) *he* (and) *yutang* (fishpond)] [P *zouqu* (went to)].

3) Body parts

Zhidao (up to) *xianzai* (now), (,) [D *wo shenshang* (on my body)] *hai* (still) [P *chuanzhe* (wear)] *ta de* (her) *yundongku* (sport pants).

By far the main attributes of *Location* come out as follows.

1). Preexisting. *Location* is independent of event, in that it neither appears with the taking place of an event, nor disappears with the termination of an event. For example:

Women (we) [D *zai* (in) *tianye* (a field)] [P *sanbu* (take a walk)].

"Tianye (Field)" in the above example is fixed, which cannot disappear with the termination of the walking.

2). Static. *Location* is a fixed place, and hence its displacement cannot happen.

Baba (father) yi (once) [P *zoujin* (walk into)] [O *woshi* (bedroom)], (,) [S *mama* (mother)] *ganjin* (quickly) *ba* (ba) *shu* (a book) [P *cang* (hide)] [C *zai* (in) *beiwo* (bed) *li* (inside)].

3). Dependent. *Location* itself cannot serve as a complete sentence, since it is an adverbial modifier after all. For example, "*cong menfeng* (from the crack between a door and its frame)" in the following sentence cannot work as a complete sentence itself.

Yiyi (Aunt) [D *cong* (from) *menfeng* (door)] [P *kuijian* (caught sight of)] *wo* (I) *bei* (was) *daoling* (turned upside down) *zhe* (zhe).

4). Marked. *Location* is marked by a prepositional marker (except for serving as subject). For example:

Shui (who) *you* (have) *zheyang* (so) *da* (big) *de* (of) *liqi* (strength) *ba* (ba) *zheme* (so) *da* (large) *kuai* (chunk) *de* (of) *shitou* (stone) [P *bandao* (move to)] [O *senlin* (forest) *li* (inside)] *lai* (to) *ne* (ne).

Location is not as prominent as other typical participants in an event structure, since it generally provides a place for an event to take place. However, *Location* is closely connected with the typical event structure. Without *Location*, the concrete context in which an event takes place is fuzzy. In this sense, *Location* is indispensable to reach the integrity of the sentence meaning.

6. **Conclusion.** In this paper, we analyzed the lexical-semantic constraints on the lexicon-to-syntax projecting. We conclude that the semantic role *Location* can be projected on adverbial, complement, object and subject syntactically, and adverbial is the most typical one. In addition, the words that belong to the C category are the main type of words that act as *Location* role.

In future work, we plan to use the annotated corpus to train an automatic semantic role labeling system, and try to explore the lexical-semantic constraints discovered in this paper to improve the overall performance.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (No. 61272215) and National Social Science Foundation of China (No. 12BYY123).

REFERENCES

[1] C. Chen, *Study on Semantic Layer of Contemporary Chinese*, Xuelin Publishing House, 2003.
[2] C. Lu, *Semantic Network of Chinese Grammar*, Commercial Press, 2001.
[3] J. Lu, On interface between syntax and semantics, *Journal of Foreign Languages*, no.5, 2006.
[4] J. Lu, We should pay more attention to study and description on features, *Yangtze River Academic*, no.1, 2006.
[5] Y. Yuan, The fineness hierarchy of semantic roles and its application in NLP, *Journal of Chinese Information Process*, no.4, 2007.
[6] D. Sun, Constructing the tagset of semantic features of syntactic categories, *Applied Linguistics*, no.2, 2012.
[7] W. Guo and Y. Wang, The concept role relationship between scene and place, *Journal of Xi'an International Studies University*, no.9, 2011.
[8] J. Mei, *Tongyici Cilin*, Shanghai Lexicographical Publishing House, 1984.
[9] J. Shao, On bi-directional semantic selection principle of Chinese grammar, *Journal of Chinese Language*, no.8, 1997.
[10] X. Lin and C. Lu, Subjective and objective information of Chinese sentences on the semantic dimension, *Chinese Language Learning*, no.5, pp.8-11, 1997.
[11] C. Lu, *Parataxis Network of Chinese Grammar*, Commercial Press, Beijing, 2001.
[12] F. Chen, B. Cai, K. Chen and J. Huang, The construction of Sinica Treebank, *Computational Linguistics and Chinese Language Processing*, vol.4, no.2, pp.87-104, 1999.
[13] Z. Dong and Q. Dong, *HowNet and the Computation of Meaning*, World Scientific Publishing Co., Inc., 2006.
[14] Z. Chu and J. Chu, On location-role object and the use or disuse of its attribute mark, *Studies in Language and Linguistics*, vol.26, no.4, pp.89-93, 2006.