John Benjamins Publishing Company

The *bei* passive and its discourse motivations

Feng-hsi Liu*
University of Arizona

This study examines the variation between the active and the *bei* passive in Mandarin Chinese from a probabilistic perspective. The variables considered include discourse continuity factors and adversity. Two different models were built for the active-agentless passive variation and the active-agentive passive variation. Four factors were found to have significant effect: agent thematicity, patient thematicity, adversity, and referential distance. In contrast, the effect of topic persistence and local environment is not significant. The accuracy of prediction for the active-agentless passive variation is significantly higher than the accuracy for the active-agentive passive variation. Overall, the *bei* passive, either agentless or agentive, is more likely to be chosen over its active counterpart, if it is adversative, has a non-thematic agent, a thematic patient, and a shorter referential distance for the patient.

Keywords: *bei* passive, probability, variation, thematicity, adversity

关键词：被字句，概率，变换，主题性，逆境

1. Introduction

The *bei* passive is the major passive construction in Mandarin Chinese. It is also one of the better-studied constructions in Mandarin. Much research has been carried out on the structural properties (e.g. Hashimoto 1987, Huang 1999, Shi 1997, Tang 2001, Ting 1998), semantic properties (e.g. Li 1980, Li & Thompson 1981, Chappell 1986), comparisons with English (e.g. H. Wang 1983, Xiao et al. 2006) and historical development (e.g. Wang 1957, 1958, Zhang 1991) of the *bei* passive, but few studies, except for Xing (1993) and Myhill & Xing (1994), have considered its usage in discourse. In this study I will conduct a variation study on when the *bei* passive is more likely to be chosen over its active counterpart. The factors to be examined are of two categories: discourse continuity and adversity; the former has been shown to play an important role in the use of passive cross-linguistically (e.g.

Cooreman 1984, Thompson 1987, Rude 1988, Xing 1993, Myhill & Xing 1994, Biber et al 1999), while the latter has been identified as a characteristic of the *bei* passive (Wang 1945:128, Wang 1958:436, H. Wang 1957:57, Chao 1968:703; Li & Thompson 1981:493, Chappell 1986). The main issues to be addressed are the following: When is the *bei* passive more likely to be chosen over its active counterpart under the two discourse contexts? Which factors are good predictors of the variation?

The *bei* passive describes an event where the patient has no control over the event.[1] This can be taken as a central characteristic of the construction. As we will see later, this characteristic imposes restrictions on the availability of the construction in a number of ways, including the types of verbs and modals that are used in the construction. There are two types of *bei* passives, distinguished by whether the agent is expressed: the short agentless passive and the long agentive passive, as illustrated in (1):

(1) a. 因为技术不好，每道菜都烧焦了，自己还被烫到.

Yinwei jishu buhao, mei dao cai dou shaojiao-le, ziji hai
because skill not-good every CL dish all burned-PERF self also

bei tangdao
BEI burned-RES

'Because my skill was not good, every dish was burned, and I myself also got burned.'

b. 他常为了上课看漫画书被老师处罚.

Ta chang weile shangke kan manhua shu bei laoshi chufa
he often because in-class read comics-books BEI teacher punish

'He was often punished by the teacher for reading comics books in class.'

(1a) is an example of the short passive, where the agent is not expressed, and (1b) is an example of the long passive, and the agent is expressed. Note that in both types of the passive the word *bei* is present.

The *bei* passive is often thought of as a construction that describes an unhappy or unfortunate event for the patient (Wang 1945, Wang 1958:436, H. Wang 1957:57, Chao 1968:703; Li & Thompson 1981:493, Chappell 1986). But it is also noted that in Modern Chinese, due to Western influence, there has been an increase of *bei* sentences that express non-adversative situations. In this study, adversity will be included as one of the factors to be investigated.

In previous studies, a number of discourse factors have been proposed that favor the use of the passive cross-linguistically, including topic continuity (Cooreman et al 1984, Rude 1988, Xing 1993, Myhill & Xing 1994), thematic continuity

(Thompson 1987), information status (Birner & Ward 1998), and immediately surrounding clauses (Thompson 1987). However, there has not been much research that specifically looks at the *bei* passive from a discourse perspective. Xing (1993) and Myhill & Xing (1994) are among the few studies in this direction. The two studies are conducted in the tradition of Givón (1983), where actives and passives are compared in terms of topic continuity. Two dimensions of topicality are used — referential distance (RD) and topic persistence (TP). RD measures the distance between the agent (or the patient) and its closest antecedent, while TP measures the persistence of the agent (or the patient) in following clauses. Both measures are based on number of clauses. In the passive, rather than the active, the agent is shown to have a higher mean value of RD than the patient, and the patient has a higher mean value of TP than the agent. These results are found in Xing's (1993) study.

But a comparative study does not quite answer the question of when the *bei* passive is likely to be used. In fact, the findings based on comparisons of means of RD and TP have never been put to test. Are they good predictors of the use of the passive? Furthermore, the variation between the active and the passive may also be sensitive to factors other than topic continuity. What are some of the other factors? And how much role does each factor play? Recent studies (e.g. Bresnan 2007, Bresnan & Ford 2010, Yao & Liu 2010) have shown that a good way to examine variation is a probabilistic approach where the choice of a variant is predicted by a range of factors. This is the approach I will adopt. In the active-*bei* passive variation, the likelihood of a sentence being used in either the active form or the passive form is subject to semantic, pragmatic and discourse factors. In this study, besides adversity, I will consider three discourse properties all having to do with continuity: topic continuity, thematic continuity, and continuity in the immediate surrounding environment. All three properties have been proposed in the literature as factors contributing to the use of passives in general, as mentioned above. Taking the continuity factors as well as adversity as variables, my goal is to find out what effect they have on the variation between the active and the passive. I will attempt to answer two questions: (a) How successful are the factors in predicting the active-passive variation? (b) Which of the factors are significant? As a variation study, it will include both passive and active data. In addition, given that there are two types of *bei* passives, one would like to know whether these discourse factors affect their choice in a similar way. I will therefore consider the two types of passives separately.

Before presenting the data and methods of measurement, I will consider the environments where the *bei* passive can and cannot be used. To understand the active-passive variation, it is first necessary to understand the context under which the variation is possible.

2.   Environments for the *bei* passive

Unlike English, where most transitive active clauses can be turned into a passive form (Wang 1957, 1958), in Mandarin Chinese the *bei* passive is not a productive construction, as it is only available with certain verbs and in limited environments. Typologically, Mandarin Chinese can be considered to be low on the scale of productivity of passive, although it is not at the bottom, as certain languages do not allow passive at all, such as Chadic languages and many languages in New Guinea (Keenan 1985). Li & Thompson (1976) take the low-productivity of *bei* passive as a typological characteristic of topic-prominent languages. The limitation of the construction is mainly due to the meaning of the passive marker *bei,* which is derived from the verb *bei* 'suffer'. In earlier times *bei* was only associated with events which are unfortunate for the patient. In Modern Mandarin, however, *bei* has been extended and is used for non-adversative events as well, although this is generally seen in the written language; in spoken language, *bei* remains mostly negative (Wang 1958, Li & Thompson 1981, Chappell 1986, Xiao et al 2006). In general, three groups of verbs occur in the *bei* passive, reflecting three stages of development of the construction.

The first group consists of verbs that have the adversative meaning, as illustrated in (2):

(2)   Verbs which imply adversity:
骂 *ma* 'scold', 咬 *yao* 'bite', 处罚 *chufa* 'punish', 赶出去 *gan chuqu* 'chase out', 绑架 *bangjia* 'kidnap', 革职 *gezhi* 'dismiss', 取笑 *quxiao* 'laugh at', 抢走 *qiangzou* 'rob, taken away', 开除 *kaichu* 'expel', 虐待 nuedai 'abuse', 淘汰 *taotai* 'eliminate'

Examples of verbs of negative coloring which occur with the *bei* passive can be seen in (1) above. This could be considered as the core use of the *bei* passive. In such sentences the verbs are adversative and *bei* is directly related to the original meaning of 'suffer'.

The second group consists of verbs that are not adversative, but the event described acquires the adversative quality when occurring with *bei*. This is illustrated in (3):

(3)   Verbs where the negative meaning is absent or weak, but often becomes adversative when occurring with *bei*:
看见 *kanjian* 'see', 知道 *zhidao* 'know', 记下名字 *jixia mingzi* 'record the name', 碰见 *pengjian* 'run into', 定型 *dingxing* 'set into mold', 发现 *faxian* 'find out'

As noted by Li & Thompson (1981: 495–496), Chappell (1986), and Hashimoto (1987), a *bei* passive with *kanjian* 'see' has the added meaning that the event of being seen is unfortunate, illustrated in (4a):

(4)   a.   张三被人看见了 (Li & Thompson 1981: 496)

Zhangsan bei ren   kanjian le
Zhangsan BEI person see   PERF
'Zhangsan was seen by people.'

b.   我看见了张三

Wo kanjian -le   Zhangsan
I   see   -PERF Zhangsan
'I saw Zhangsan.'

On the other hand, (4b) carries no such implications. The negative meaning can be attributed to the meaning of the construction as a whole, rather than the meaning of the verb.

In the third group, neither the verb nor the construction implies adversity. These verbs are further divided into two sub-groups. The first sub-group includes compound verbs of the form X-*wei* 'X-as', while the second sub-group are disyllabic verbs, illustrated in (5):

(5)   Verbs that do not have the adversative meaning:
a.   compound verbs of the form X+ *wei*: 认为 *renwei* 'regard as', 称为 *chengwei* 'refer to as', 形容为 *xingrong wei* 'describe as', 聘为 *pin wei* 'hire as', 视为 *shi wei* 'consider as', 指定为 *zhiding wei* 'assign as'
b.   鼓励 *guli* 'encourage', 领养 *lingyang* 'adopt', 安置 *anzhi* 'place', 录取 *luqu* 'admit/accept', 提名 *timing* 'nominate', 核准 *hezhun* 'approve', 告知 *gaozhi* 'tell', 问到 *wendao* 'ask', 吸收 *xishou* 'absorb', 接受 *jieshou* 'accept', 提起 *tiqi* 'mention'

These verbs are mostly *shumian yu* 'written language'. They came to be associated with *bei* through influence from European languages (Wang 1958, Chao 1968, Li & Thompson 1981, Chappell 1986). This is especially apparent in the X-*wei* 'X-as' verbs, since the English counterpart of these verbs frequently occur in the passive. Sentences with such verbs represent a more recent development of the construction, as in (6):

(6)   a.   他被认为是长人队最突出的投手

Ta bei renwei   shi Changren dui   zui   tuchu
he BEI think-of is   Changren team most outstanding

de toushou
DE pitcher

'He was thought of as the most outstanding pitcher on the Changren team.'

b. 我们递出的申请被核准了

Women dichu　　de shenqing　bei hezhun le
we　　send-out DE application BEI approve PRT
'The application we sent out was approved.'

We can also list some of the environments that are incompatible with the *bei* passive. First, a range of verb types do not occur with *bei*, as illustrated in (7):

(7) Verbs that usually do not occur with *bei*:
  a. Verbs that are non-adversative, used in spoken language, and with an inanimate patient: 做好 *zuohao* 'make-finish', 洗干净 *xiganjing* 'wash it clean', 写完 *xiewan* 'write-finish', 说清楚 *shuo qingchu* 'say it clearly'
  b. verbs whose agent, if expressed, is marked by 由 *you*, which indicates a strong sense of agentivity: 负责 *fuze* 'bear responsibility', 主持 *zhuchi* 'in charge of (a meeting)', 扮演 *banyan* 'play (a role)', 提供 *tigong* 'provide', 出版 *chuban* 'publish', 举办 *juban* 'host (a conference)'
  c. light verbs:[2] 进行 *jinxing* 'do', 加以 *jiayi* 'do something to'
  d. verbs that take clausal complements: 希望 *xiwang* 'hope', 觉得 *juede* 'think'

This list is not meant to be exhaustive; even so, it shows that a number of factors are at work in restricting the types of verbs in the *bei* passive, including animacy, agentivity, and types of complements. Note that the verbs listed in (7) are not absolutely barred from the *bei* passive; it's just that in corpus data, e.g. the Sinica Corpus, these verbs are usually not found in a *bei* passive sentence. Even if a verb can occur with *bei*, the structure it occurs in is often limited. For example, even though *zhidao* 'know, find out' in the active voice can freely take nominal as well as sentential objects, as in (8), in a passive *zhidao* is much more comfortable taking a nominal subject, as in (9a), where the passive clause serves as complement of the verb *pa* 'afraid'. When *zhidao* takes a sentential subject, the result is less acceptable, as (9b) shows.

(8) a. 我早已经知道这件事了
      Wo zao　yijing　zhidao zhejian shi　　le
      I　early already know　this-CL matter PRT
      'I already knew about this matter early on.'

    b.   大家都知道小李没考上大学

        Dajia   dou zhidao Xiaoli mei   kaoshang daxue

        everyone all   know   Xiaoli not-PERF get-into   college

        'Everyone knows that Xiaoli didn't get into college.'

(9) a.   他怕这件事会被别人知道

        Ta pa   zhejian shi   hui bei bieren   zhidao

        he afraid this-CL matter will BEI other-people know

        'He's afraid this matter will be found out by other people.'

    b.   [?]他怕去年没考上大学会被被人知道

        [?]ta pa   qunian   mei   kaoshang daxue hui bei

        he afraid last-year not-PERF get-into   college will BEI

        bieren   zhidao

        other-people know

        [?]'He is afraid that he did not get into college will be found out by other people.'

Besides verbs, there are also other factors that affect the availability of *bei*. For example, modality plays a role. The modal *hui* 'will' occurs with the *bei* passive, but the modal *yao* 'will' is less likely to, although it occurs with the *bei* passive when accompanied by *kuai* as in *kuaiyao* 'about to', or negation, *bu yao* 'do not':

(10) 你这样做会被批评的

     Ni zheyang zuo hui bei piping   de

     you this-way do   will BEI criticize PRT

     'If you do things this way you will be criticized.'

(11) a.   小李可能 ?要/会 被审问

        Xiaoli keneng yao/hui bei shenwen

        Xiaoli possible will/will BEI interrogate

        'It is possible that Xiaoli will be interrogated.'

    b.   小李快要被审问了

        Xiaoli kuaiyao bei shenwen   le

        Xiaoli about-to BEI interrogate PRT

        'Xiaoli is about to be interrogated.'

    c.   你千万不要被他骗了

        Ni qianwan bu yao bei ta   pian   le

        you definitely not will BEI him fooled PRT

        'Be sure not to be fooled by him.'

This restriction is a manifestation of how the concept of control regulates the *bei* passive. Since in the construction the patient has no control over the event being

described, *yao* 'will', when used in the volitional sense, is incompatible in this environment. But *kuai yao* 'about to' in (11b) is non-volitional, and in (11c) the volition of *bu yao* is with the speaker, rather than the addressee (the subject).

Another factor that plays a role is telicity. This is illustrated in (12):

(12) a. ᵀ小李被小张喜欢了
          Xiaoli bei Xiaozhang xihuan le
          Xiaoli BEI Xiaozhang like   PERF
          'Xiaoli was liked by Xiaozhang.'
      b. 小李被小张喜欢上了
          Xiaoli bei Xiaozhang xihuan shang le
          Xiaoli BEI Xiaozhang like   up   PERF
          'Xiaoli was liked by Xiaozhang (Xiaozhang got interested in Xiaoli).'

(12a) shows that *xihua*n 'like' does not occur with *bei* comfortably, but when *shang* 'up', which turns an atelic verb into telic, combines with *xihuan*, the resulting verb is compatible with *bei*, as in (12b). Thus when a verb is telic, it increases its chances of occurring with *bei*.[3]

In short, in the above I have shown that the issue concerning availability of the *bei* passive is a complicated one, having to do with the verb, aspect, animacy, agentivity, telicity and the concept of control. All of this suggests that active and passive sentences are not parallel; there is an asymmetry between the two: the active form is the basic form, and is generally available, while the *bei* passive is a marked form and is available only under certain circumstances.

3. This study

Having identified some of the restrictions on the *bei* passive, in this section I turn to the issue that is the focus of my study, namely, when is the *bei* passive more likely to be chosen over its active counterpart? Just because a *bei* passive can be used does not mean that it will be used in a given context. A number of factors may be responsible for a speaker's choice of the *bei* passive over the active. I will use three methods to measure textual cohesion: thematic continuity (thematicity), topic continuity (topicality) and continuity in immediately preceding environments. Below in 3.1 I introduce the data that is used for the study, and in 3.2 I discuss the methods for the three measures.

3.1 Data

The data comes from the Sinica Corpus, which contains 5 million characters of Modern Chinese. I collected both active and passive clauses. With the *bei* passive, I first extracted every tenth clause in the Sinica Corpus that occurs with *bei*. Four types of data were then excluded because they are either not instances of *bei* passives, or they are *bei* passives used as modifiers. My focus is *bei* passives that have a non-modifying function.

First, lexical passives were excluded, e.g. *bei po* 'be forced'. In a lexical passive, *bei* forms a compound with the following monosyllabic verb, which otherwise cannot function as an independent verb in Modern Chinese. Another type of data that was excluded are compound nominals that include *bei,* e.g. *beigao* 'the defendant' (literally, the one that was accused), *beihairen* 'the victim' (literally, the person that was victimized). The third type of sentences excluded is relative clauses that are in the form of *bei* passive, as in (13):

(13) 但却一直没法提出适当而且可被居民接受的补偿条件

　… dan que yizhi 　　 meifa 　 tichu shidang 　　 er
　　 but but continuously not-able offer appropriate and

　ke 　 bei jumin 　 jieshou de buchang 　　 tiaojian
　able BEI residents accept 　 DE compensate terms

　'…but all this time (they) were not able to offer appropriate compensations that can be accepted by the residents.'

The passive clause *ke bei jumin jieshou* 'can be accepted by residents' is a relative clause, and such data is not included in the database. The fourth type of data also includes the *bei* passive functioning as a modifier, as in (14):

(14) 绝大多数的海豚至今仍面临被屠杀的命运

　Juedabufen de haitun 　 zhijin 　　 reng mianlin bei
　most 　　　 DE dolphin to-today still 　 face 　　 BEI

　tusha de mingyun
　kill 　 DE fate

　'Most of the dolphins today still face the fate of being killed.'

The clause *bei tusha* 'be killed' acts as a modifier of *mingyun* 'fate'. Such clauses are also excluded.

The total number of *bei* passive clauses that resulted is 549.

Besides data of the *bei* passive, I also extracted a sample of active clauses in the corpus to be included in the statistical analysis. The selection of tokens was based on one criterion: the active clause has the potential to be expressed in the form of the *bei* passive when considered in isolation. That is, the verbs, modals or aspect in these clauses are also compatible with the passive use. This will eliminate lexical, semantic or syntactic incompatibility as possible reasons for not using the passive. The following is an example that meets the criterion:

(15)   警方只有在现场查获一千二百元零钱, 并捉到来不及逃走的上述四人及
碗公和骰子, <u>随即警方将四人带回处理</u>.

Jingfang zhiyou zai xianchang chahuo yiqian erbai yuan
police   only   at on-site   find   1000   200   CL

lingqian,   bing zhuodao laibuji   taozou   de
loose-money also catch   lack-of-time run-away DE

shangshu   si   ren   ji   wangong he   shaizi,
above-mentioned four people and bowl   and dice

<u>suiji   jingfang jiang si   ren   dai hui   chuli</u>
immediately police   OBJ   four people take back handle

'The police only found 1200 yuan at the site; they also arrested the above-mentioned four people who weren't able to get away in time, (picked up) some bowls and dice. <u>The police immediately took them back (to the police station)</u>.'

The underlined clause can also be expressed as a *bei* passive, with or without the agent: *si ren suiji bei (jingfang) daihui chuli* 'the four people were immediately brought back (by the police).' Notice that it is not required that the alternative *bei* passive be used in the context; I am only concerned with whether an active clause has a *bei* passive form. I first identified two most frequent types of verbs that occur with the *bei* passive, VC (active transitive verbs) and VG (classificatory verbs),[4] in the corpus. I first extracted every tenth active clause that contained these two types of verbs. I then went over each sentence and decided if a given active clause has the potential to be expressed in the *bei* passive. 660 tokens met this condition, and half, chosen from every other token in the list, were included in the database.

Thus the data includes 330 passages with an active clause and 549 passages with a *bei* passive clause, of which 348 are agentless and 201 are agentive. Altogether 879 passages form the database for my analysis.

3.2 Factors

This section discusses how each of the three factors was coded.

Thematicity is concerned with continuation of the same theme throughout a paragraph or a section of the discourse. Givón (1983) discusses three types of discourse continuity: thematic, action and topic. He considers thematic continuity the 'overall matrix' of all continuity. Givón also says it is the least specified and least coded in structure. Thompson (1987) proposes thematic continuity as a factor that favors the agentive passive over the active in English. According to Thompson, if the patient is more related to the theme of the paragraph than the agent, then it counts as having more thematic continuity than the agent, and in this situation, a passive clause, rather than an active clause, is more likely to be used. In this study two features were coded with respect to thematicity: agent thematicity, and patient thematicity. These are binary features, depending on whether the agent/patient is thematic or not. Two examples are provided below:

(16)   记得小时候，常做一个梦，梦见<u>自己被关在黑暗的地方</u>，全看不见东
西，只能靠手来猜测碰到的物体，我在空间中游走，不断地眨着眼
皮，仍是一片漆黑.

Jide   xiao shihou, chang zuo   yige   meng,
remember small time   often dream one-CL dream

mengjian <u>ziji bei guan zai heian de difang</u>, quan
dream-of self BEI lock at dark DE place   all

kan-bu-jian dongxi, zhineng kao   shou lai caice
look-not-see things   only-can depend-on hand DIR guess

pengdao de wuti, wo zai kongjian   zhong youzou,
touch   DE things I   at empty-space in   move-about

buduande   zhazhe   yanpi, rengshi yipian qihei
continuously blink-DUR eyelids still   one-CL darkness

'I remember when I was little, I had a recurrent dream. I dreamed of <u>myself being locked in a dark place</u>, not being able to see a thing, and I could only use my hands to guess what it was that I touched. I moved about aimlessly, continuously blinking my eyes, but it was still dark all over.'

The target clause is underlined. In this example, the patient, referring to the speaker herself, is thematic, while the agent, which is not expressed, is not thematic, as it is irrelevant to the story line. (17) is another example:

(17) 我也听到些主任叙述自己处里的组长 年轻气盛，我行我素，不但工作
无法顺利推展，也得罪许多人，连自己也被殃及，一天到晚忙着替他「
处理善后」。还有那完全被动型的，得每天耳提面命，叮咛一桩做一
桩。如果哪件事，主任忘了提醒，组长绝不会「偶然想起」。

Wo ye   tingdao xie   zhuren xushu   ziji chuli de
I   also heard   some directors describe own bureau DE

zuzhang   nianqingqisheng woxingwosu,   budan
section-head young-energetic do-one's-own-way not-only

gongzuo wufa   shunli   tuizhan, ye   dezui xuduo ren,
work   cannot smoothly propagate also offend many   people

lian ziji ye   bei yangji, yitiandaowan mang-zhe ti ta
even self also BEI affect   all-day   busy-DUR for him

chuli   shanhou. haiyou na   wanquan beidong xing de,
take-care-of aftermath also   that complete passive type DE

dei   meitian   ertimianming, dingning yi   zhuang zuo yi
have-to everyday remind   remind   one CL   do one

zhuang. ruguo najian   shi   zhuren wang-le   tixing,
CL   if   some-CL matter director forget-PERF remind

zuzhang   jue   bu hui ouran   xiangqi
section-head definitely not will incidentally think-of

'I have also heard of directors describing the section heads in their bureau as being young, energetic, and doing things in their own way. Not only is the work difficult to carry out, they also offend many people. Even the directors themselves are affected. All day long they (the directors) are busy taking care of the aftermath for them (the section heads). And as for ones that are completely passive, the directors need to remind them every day. If you remind them of one thing, they do one thing. If there is something that the director forgets to remind them of, the section head would never think of it incidentally.'

In this passage, both the agent, referring to the section heads, and the patient, referring to the directors, are thematic, as the passage is concerned with both.

Next, we consider topic continuity, which is introduced in Givón (1983). As mentioned earlier, studies of the passive that discuss topic continuity usually consider two dimensions: referential distance (RD) and topic persistence (TP). I will follow this practice. On RD, I measure the distance between the agent (and the patient) and its antecedent. If the antecedent of the agent or the patient occurs in

the immediately preceding clause, a value of 1 is assigned. In Givón (1983), the domain of referential distance is set to be 20 clauses: however, later (Givón 1994) only three values are included: 1, 2/3 (when RD value is 2 or 3) and 3> (when RD value is more than 3). In this study I will set the domain of RD to be 10 clauses. If the agent or patient does not occur within the 10 clause domain, a value of 11 is assigned. When the RD domain is less than 10, which happens when the target clause occurs close to the beginning of a passage, the range of the values goes from 1 to the number of clauses preceding the target clause plus 1. For example, if the target clause is the 4th clause, the range of RD goes from 1 to 4.

There are two ways to measure TP (Myhill 1992): by persistence, in terms of the number of consecutive following clauses that contain reference to the target NP, or by frequency, in terms of how many clauses in the range of 10 following clauses that contain reference to the target NP. I will adopt the second method, taking a frequency count of a referent. The main reason is that there are a large number of tokens in the data where the patient is referred to more than once, but non-continuously, in following clauses. A frequency account is better able to capture the fact that the referent has not decayed in the domain.

To illustrate the measurement of RD and TP, we consider (16–17) again. In (16), the target clause is the fourth clause of the passage. The RD value for the patient is 1, as it is mentioned in the immediately preceding clause, while the RD for the agent it is 4. which is not mentioned in any of the preceding three clauses. For TP, in this stretch there are six clauses following the target clause, including the relative clause.[5] The patient occurs in five of them and has a TP value of 5, while the agent occurs in none of them and has a TP value of 0. In (17), the RD value for the patient, referring to the director, is 4, as it is mentioned in the fourth clause counting backwards from the target clause,[6] while the agent, referring to the section head, has an RD value of 1. As for TP, in the stretch here, there are six following clauses,[7] both the patient and the agent are mentioned in four of the six clauses, and both have a TP value of 4.

RD and TP both serve as variables. However, rather than taking the mean value of the agent and the patient, as is done traditionally, I take the difference value as the value of each variable. For both RD and TP, I take the patient-agent difference as the value. Thus in (16) the RD difference is -3 (1–4), while the TP difference is 5 (5–0), and in (17) the RD difference is 3 (4–1), while the TP difference is 0 (4–4). Relative topicality of each token captures the relationship between the agent and the patient on individual basis. Givón (1994), following Cooreman (1987), defines voice patterns in terms of relative topicality between the agent and the patient. Passive is pragmatically defined as a structure where the agent is much less topical than the patient. This measurement follows Givón's definition of the passive.

The third type of continuity to be measured is the immediate environment, in particular, the immediately preceding clause of the target clause. Both the agent and the patient will be coded as to whether either one is mentioned in the immediately preceding clause. Thompson (1987) takes the immediately preceding clause as a factor for the use of the passive in English. If the patient is mentioned in the immediately preceding clause, but not the agent, then the passive is likely to be used. Myhill & Xing (1994) suggest that the immediately surrounding discourse, both before and after the target clause, which they refer to as local topicality, matters more than textual topicality in Chinese with respect to the choice between active and a passive-like construction of the form patient + verb, e.g. *chezi xiu- le* 'The car was fixed'. I include local environment as a measure in order to find out if it also affects the choice of *bei* passive. In particular, I would like to find out how local environment compares with broader textual topicality in predicting the variation.

Another feature that is coded is adversity or affectedness. A clause is either adversative or non-adversative. A neutral verb that occurs with *bei*, such as *bei kanjian*, is coded as adversative. On the other hand, *kanjian* in an active sentence is coded as non-adversative. An X-*wei* verb that occurs with *bei*, such as *bei renwei* 'is thought of as' is coded as non-adversative.

Altogether six variables serve as predictors of the variations. Four are categorical: agent thematicity, patient thematicity, local environment, and adversity, while the other two are numerical: RD difference and TP difference. Two logistic regression models were built to fit the data, the agentless model which includes active data and agentless passive data, and the agentive model, which includes active data and agentive passive data. The agentless model includes 678 tokens, of which 330 are active and 348 are passive; the agentive model includes 531 tokens, of which 330 are active and 201 are passive.

4. Results

4.1 Agentless passives

Results of a preliminary analysis showed that of the six variables, TP and local environment are not significant, as Table 1:

Table 1. Results from the preliminary analysis with 6 factors

predictor	coefficient	p value
agent thematicity	−2.64	<.0001
patient thematicity	1.70	<.0001
RD difference	−0.10	<.0001
TP difference	0.03	.3615
local environment	−0.14	.5865
adversity	1.13	.0001

Therefore, a revised model was built which included four variables, excluding TP and local environment. The results are given in Table 2:

Table 2. Results from the agentless model

predictor	coefficient	p value
agent thematicity	−2.74	<.0001
patient thematicity	1.76	<.0001
RD difference	−0.07	<.0001
adversity	1.13	<.0001
agent thematicity x RD difference	−0.09	.0336

All four variables are significant; in addition, the interaction between agent thematicity and RD difference is significant. The results show that the agentless *bei* passive is more likely to be used when the clause describes an adversative event, the agent is not thematic, the patient is thematic, and the RD value of the patient is lower than that of the agent. The values of the coefficients indicate that agent thematicity is the strongest predictor, followed by patient thematicity and adversity. RD difference is the least effective among the four predictors. If we take .5 probability as the cutoff point, the model correctly predicts 84.3% (572/678) of the data.

Below we consider some examples. (16), presented earlier, is an example that satisfies all four conditions: the agent is non-thematic, the patient is thematic, and the patient has a lower RD value than the agent, and the clause describes an adversative situation. Note that the unexpressed agent is in fact non-topical, being the only mention in the passage. (16) represents a major sub-type of the agentless passive, as in 69.0% (240/348) of the agentless passive the agent is neither topical nor thematic. (18) is another example that satisfies all four conditions:

(18) 我今天小考考八十分，回去一定会被爸爸骂死。走着走着，珮珍知道
自己到家了，却不敢进屋，她自言自语的说：回去只好老老实实的说
出来，本来就是我自己不好，没有好好看书，<u>被处罚也是应该的</u>。于
是珮珍开门进去，只见客厅没有半个人，她便在客厅等爸爸妈妈回
来。说也奇怪，平常爸妈五点回到家，现在已经五点五十分了...

Wo jintian xiaokao kao bashi fen, huiqu yiding   hui bei
I   today quiz   test 80   score return certainly will BEI

baba masi.   zou- zhe zou-zhe, Peizhen zhidao ziji
father scold-dead walk- DUR walk-DUR Peizhen know   self

daojia   le, que bugan   jinwu,   ta ziyanziyu -de
arrive-home PRT but not-dare enter-house she self-talk   -ADV

shuo: huiqu   zhihao   laolaoshishi -de shuo chulai, benlai
say   go-back can-only honest   -ADV say   out-DIR originally

jiushi woziji buhao meiyou   haohao kanshu, <u>bei chufa   yeshi</u>
it- is myself bad   not-PERF well   study   BEI punish is

<u>yinggai   de</u>. yushi Peizhen kai   men jinqu, zhijian   keting
deserved PRT then Peizhen open door enter only-see living-room

meiyou ban ge ren,   ta bian zai keting   deng   baba
no   half CL person she then at living-room wait-for father

mama   huilai. Shuo ye   qiguai, pingchang bama
mother return. say   PRT strange normally   father-mother

wudian   huidao jia, xianzai yijing wudian   wushifen   le
five-o'clock return home now   already five-o'clock fifty-minute PRT

' "I got 80 on the quiz today. When I get home, father will scold me to death
for sure." As she kept walking, Peizhen knew she was home, but she was
afraid to go in. She said to herself: "When I get in, I have to tell the truth. It
is my own fault; I didn't study hard. <u>I deserve to be punished</u>." Then Peizhen
opened the door and went in. There was no one in the living room. So she
waited in the living room for mom and dad to return. It was strange. Usually
mom and dad came home at 5:00; it was already 5:50 by then.'

In this example the agent, referring to Peizhen's father, is mentioned once before
the target sentence, and twice afterwards, but it is non-thematic, as the passage is
about Peizhen, not about Peizhen's father.

Since TP difference is not a significant variable, whether the agent or the pa-
tient has the higher value does not affect the variation. In both (16) and (18) above

the patient has a higher TP value, but (19), where the agent of the target sentence is more persistent than the patient, is also predicted to have the passive form:

(19)   又走了一阵，忽听得一个女子尖声大叫，依稀正是阿曼。苏普循声奔去推开一扇门，只见阿曼缩在屋角之中，<u>双手被反绑在背后</u>。两人惊喜交集，齐声叫了出来。苏普抢上去松开了她的绑缚，问：「那恶鬼呢？」阿曼道：「他不是鬼，是人。刚才他还在这里，听到你们的声音，便想抱了我逃走…

You   zou -le   yizhen, hu   tingde yige   nüzi
more walk-PERF a-while suddenly hear   one-CL woman

jiansheng   dajiao, yixi   zhengshi Aman. Supu xunsheng
piercing-noise yell   seem just-is   Aman Supu follow-sound

benqu   tuikai   yishan men, zhijian   Aman suozai wujiao
run-over push-open one-CL door only-see Aman huddle corner

zhizhong, <u>shuangshou bei fanbang   zai beihou</u>. liangren
in   both-hands BEI backwards-tie at back   two-people

jingxi   jiaoji, qisheng jiao-le   chulai. Supu qiang
surprise-happy mixed together yell-PERF DIR   Supu rush

shangqu songkai-le tade bangfu wen: na   egui   ne? Aman
over   untie-PERF her bind   ask that ghost PRT Aman

dao: ta bu shi gui, shi ren   gangcai ta hai zai zheli
say he not is ghost is person just-now he still at   here

tingdao nimende shengyin, bian xiang bao -le   wo taozou,
hear   your   noise   then want carry -PERF me escape

'(He) walked some more. Suddenly he heard a woman screaming. It sounded like Aman. Following the sound, Supu ran over. As he opened a door, he saw Aman huddled at a corner, and <u>her hands were tied backwards behind her back</u>. The two of them, shocked and happy, yelled out loud together. Supu rushed over and untied her. He asked: "Where is that ghost?" Aman said: "He is not a ghost; he is a man. Just now he was still here. He heard your noise, and he wanted to carry me and escape…"

The unexpressed agent refers to the 'ghost', it is mentioned six more times after the target sentence, whereas the patient, referring to Aman's hand (therefore Aman), is mentioned five times in the following ten clauses. I will return to the role of TP in Section 5.

When the agent is thematic, however, a clause is predicted to occur in the active. (20–21) are two examples:

(20)   我觉得我要是这节课没有上到，我可能落后别人很多，<u>漏掉很多东西</u>，那我没有办法负担这个后果，所以我不敢翘课。

Wo juede wo yaoshi zhe jie ke   meiyou shangdao, wo
I   feel   I   if   this CL class not-PERF go-RES   I

keneng luohou   bieren   henduo, <u>loudiao henduo dongxi</u>,
may   fall-behind other-people a-lot   miss   a-lot   stuff

na   wo meiyou banfa fudan   zhege   houguo,
then I   not-have ways be-responsible this-CL consequence

suoyi   wo bugan   qiaoke
therefore I   dare-not cut-class

'I feel if I didn't go to a class, I may fall behind other people a lot and <u>miss a lot of things</u>. Then I have no way to take the consequences. Therefore, I dare not cut classes.'

(21)   由于每年必须支付的贷款利息高达新台币三、四千万元，<u>几乎把金朋正常营业所赚取的利润消耗殆尽</u>，因此这些年我们大多是处于损益两平的情势。事实上，为了搏节开支，去年金朋还进行了裁员。

Youyu meinian bixu   zhifu de daikuan lixi   gaoda
because each-year have-to pay   DE loan   interest reach

xintaibi san si   qianwan   yuan, <u>jihu   ba Jinpeng</u>
NTD   three four 10-million dollar almost BA Jinpeng

<u>zhengchang yingye   suo zhuanqu de lirun xiaohao daijin,</u>
normal   business PRT earn   DE profit use   almost-gone

yinci   zhexie nian women daduo shi chuyu sunyi
therefore these   years we   mostly are at   loss-profit

liangping de qingshi. Shishishang, weile   zunjie kaizhi,
even   DE situation in-fact   in-order save   expenses

qunian   Jinpeng hai jinxing-le caiyuan
last-year Jinpeng also do-PERF   cut-staff

'Because each year the loan interest reaches 30 to 40 million New Taiwan Dollars, <u>which almost uses up the profit that Jinpeng earns in normal business</u>, these years we are mostly in a balanced situation of no loss and no profit. In fact, in order to reduce expense, last year Jinpeng also cut some staff.'

In (20) the agent, referring to the speaker, is thematic, but the patient, referring to things the speaker would miss in class if he is absent, is not thematic. In (21) both the agent, referring to the mortgage interest, and the patient, referring to the profit made by the company, are related to the theme, the financial situation of the company. In both passages the use of the active is correctly predicted.

The model also leaves 15.7% (106/678) of the tokens unaccounted for. Below is an example that is predicted by the model to be in the active, but occurs in the passive:

(22) 到工地前，他都事先打听这工地有什么老人，他往往能在见一两次面
后，就记得对方的姓名。有的话，他一到工地就叫：「老刘，怎么
样，家里的孩子怎么样？」<u>大家都被叫得心里很舒服</u>，心想：「老总
在三四千人中还记得我，我一定很重要」，也就更卖力地工作。

Dao gongdi qian, ta dou shixian dating zhe gongdi
go-to work-site before he always in-advance find-out this site

you shenme laoren, ta wangwang neng zai jian yiliang
there-is what old-people he often can at see one-two

ci mian hou jiu jide duifang de mingzi. you dehua
CL face after then remember others DE name have if

ta yi dao gongdi jiu jiao: Laoliu, zenmeyang
he as-soon-as arrive site then say Old-Liu how

jiali de haizi zenmeyang <u>dajia dou bei jiaode xinli</u>
home DE children how everyone all BEI call-RES heart-in

<u>hen shufu</u> xin xiang: Laozong zai sansi qian
very comfortable mind think old-chief in three-four thousand

ren zhong hai jide wo, wo yiding hen zhongyao, yejiu
people among still remember me I must very important then

geng maili de gongzuo
more work-hard DE work

'Before he goes to the work site, he would always try to find out whether there are old people at the site. Often he is able to remember people's names after only one or two encounters. If there are old people, as soon as he arrives at the site, he would say: "Old Liu, How is it going? How are your children?" <u>Everyone, being addressed this way, feels great inside</u>. They think: "The old chief remembers me among three, four thousand people. I must be very important." Thus they work even harder.

This passage is about a CEO who is good at remembering names of workers and cares about them. Both the agent, referring to the CEO, and the patient, referring to the workers, are thematic. In addition, adversity does not hold as the clause describes a positive situation.

In short, in the active vs. agentless passive variation, the two most important factors are agent thematicity and patient thematicity.

4.2 Agentive passives

For the agentive passives, I again performed a preliminary analysis first. Of the six variables, TP and local environment were again found to be not significant, as shown in Table 3:

Table 3. Results from the preliminary analysis with 6 factors

predictor	coefficient	*p* value
agent thematicity	−1.11	<.0001
patient thematicity	1.33	<.0001
RD difference	−0.12	<.0001
TP difference	0.04	.2313
local environment	−0.11	.6156
adversity	1.35	<.0001

Therefore, in the revised model, four variables were included: agent thematicity, patient thematicity, RD difference and adversity. The results are provided in Table 4:

Table 4. Results from the agentive model

predictor	coefficient	*p* value
agent thematicity	−1.34	<.0001
patient thematicity	1.38	<.0001
RD difference	−0.12	<.0001
adversity	1.34	<.0001
agent thematicity x RD difference	−0.10	.0105

All four variables are significant, and the interaction between agent thematicity and RD difference is also significant. It is predicted that all else being equal, the agentive passive is more likely to be used when the agent is non-thematic, the

patient is thematic, and the agent has a higher RD value than the patient, and the described event is adversative. Patient thematicity is the top predictor, followed by agent thematicity and adversity. RD difference, as in the agentless model, has a rather small effect. 78.5% (417/531) of the data is accounted for when .5 probability is taken as the cut off point. The accuracy rate is less than the rate of the agentless model (84.3%, 572/678), and the difference is significant ($p = 0.0106$, two-tailed). This means that the two variations, active-agentless passive and active-agentive passive, behave differently with respect to the four factors. These factors are significantly less effective in the active-agentive passive variation.

An example of correctly predicted passive use is given in (23):

(23)　…我看叫你皮包骨会更恰当一点。」
「还说呢！大胖猪，我倒建议你赶快减肥，要不然哪一天胖得走不动了，大概也没人背得动你。」
「你呀！什么都不吃，先担心自己会不会被<u>飓风刮走</u>比较要紧吧!」
「我什么都不吃，我只吃又香又嫩又滑的香蕉，哪像你，什么都吃，毫无选择，多没格调。」

…"Wo kan jiao ni　pibaogu　　　　hui geng qiadang　　yidian."
　　I　　see　call you skin-wrap-bone will more appropriate some

"Hai shuo ne! Da pang zhu! Wo dao　jianyi　　ni　gankuai
　still talk　PRT big fat　　pig　I　　EMP suggest you quickly

jianfei,　　　yaoburan　na　　　yitian pang de　zoubudong　　le,
lose-weight　otherwise　which day　　fat　RES walk-not-able　PRT

dagai　　　ye　mei ren　　　beidedong ni."
probably　also no　person carry-able　you

"Ni　ya! Shenme dou bu　chi, xian danxin <u>ziji hui-bu-hui　　bei</u>
　you　PRT any　　all　not eat　first worry　self will-not-will　BEI

<u>taifeng guazou</u>　　bijiao yaojin　　　ba!"
typhoon blow-away　more　important　PRT

"Wo shenme dou bu　chi, wo zhi　chi you xiang　　you nen
　I　　any　　all　not eat I　　only eat also fragrant also tender

you hua　　　de xiangjiao, na　　xiang　　ni, shenme dou
also smooth DE banana　　where resemble you any　　all

chi, haowu xuanze,　duo　mei gediao"
eat　no　　　selection how no　taste

'"…I think calling you 'a bag of bones' is more appropriate."

"You should talk! Big Fat Pig! I suggest you lose weight quickly; otherwise one day you may not even be able to walk. Probably no one would be able to carry you either."

"You, you don't eat anything. You should first worry <u>whether you would be blown away by a typhoon</u>. That would probably be more important."

"I don't eat anything; I only eat bananas, which are fragrant, tender and smooth. I'm not like you, eating anything unselectively. What a bad taste that is!"'

This is a dialogue between a boy who is skinny and a boy who is overweight, and the patient in the *bei* passive *ziji* 'self' refers to the skinny one. The patient is thematic, while the agent, *taifeng* 'typhoon' is not, as it is not related to the theme of the dialogue — a person's weight and eating habit. In addition, the agent has no topicality at all, while the patient is mentioned in the immediately preceding clause. Thus all four conditions — non-thematic agent, thematic patient, higher RD on agent, adversity — are satisfied. Such examples, where the agent is neither thematic nor topical, constitute 36.8% (74/201) of the agentive passive. Another example of correctly predicted passive is given in (24):

(24) 有一对父子，带了一头驴子到市场去卖。<u>路途上，被一个路人看到了</u>，嘲笑他们有驴不骑，真是傻瓜。于是爸爸骑上驴背，结果被一位妇人看到，责备爸爸不爱孩子；这次换儿子骑，结果又被一位老人家骂不孝；这对父子只好一起骑驴，却又被人骂不爱护牲畜，…

You   yidui fuzi   dai-le   yitou luzi   dao shichang
there-is a-pair father-son take-PERF one-CL donkey to   market

qu mai. <u>Lutu shang bei yige   luren   kandao-le</u> chaoxiao
to sell way on   BEI one-CL passer-by see-PERF   laugh-at

tamen you lu   bu qi   zhen shi   shagua. Yushi baba
them have donkey not ride really arek fool   then   father

qishang lubei,   jieguo bei yiwei   furen   kandao, zebei
get-on donkey-back result BEI one-CL woman see   criticize

baba bu ai   haizi. Zheci   huan   erzi qi, jieguo you
father not love child this-time change-to son ride result again

bei yiwei   laorenjia ma   bu xiao. Zhedui fuzi   zhihao
BEI one-CL old-man scold not filial this-pair father-son can-only

yiqi   qi lu,   que you bei ren   ma bu aihu   shengchu…
together ride donkey but again BEI person scold not care-for animals

'A father and son took a donkey to the market to put it for sale. <u>On the way, they were seen by a passer-by</u>, who laughed at them not riding on the donkey. What fools! So the father got on the donkey, but they were seen by a woman, who blamed the father for not caring about his son. And the son rode on the donkey, but he was scolded by an old man for not respecting his father. The father and son could only ride the donkey together, but they were criticized for not caring for the animal.'

This is a story from Aesop's Fables. The target clause describes an adversative event, even though the verb *kandao* 'see' is neutral by itself. The patient, referring to the father-son pair, concerns the main characters of the story, while the agent, referring to a passer-by, is not. In addition, the agent does not have an antecedent, so it has a higher value of RD than the patient.

When both the agent and the patient are thematic, a clause is predicted to occur in the active, as illustrated in (25):

(25) 这张照片是在不合法的状况下拍摄的；这一点很明显，因为在市街上我们是不被容许拍照的，即使是利用行驶中的车辆窗户亦然。<u>结果有个「执权者」把我们拦下来申诫了一番</u>，然后一路驱车尾随我们的公车到诺格拉，监视我们。

Zhezhang zhaopian shi zai buhefa de zhuangkuang xia   paishe
this-CL   photo   is at illegal DE situation   under take

de; zhe yidian   hen mingxian, yinwei zai shijie shang women
DE this one-point very obvious   because at street on   we

shi bu bei rongxu paizhao   de jishishi liyong xingshi zhong
is not BEI allow take-picture DE even   use   drive   in

de cheliang chuanghu yiran. jieguo   you   ge zhiquanzhe
DE car   window   also as-a-result there-is CL law-enforcer

ba women lan   xialai shenjie -le   yifan, ranhou   yilu
BA us   block DIR   scold -PERF one-CL afterwards along-way

quche   weisui womende gongche dao Nuogela   jianshi women
drive-car follow our   bus   to Novgorod watch us

'This picture was taken under an illegal situation. This is very obvious, because on the street we were not allowed to take pictures, not even when using the window of a vehicle in motion. <u>As a result, a 'law-enforcer' blocked us and scolded us</u>; then he followed our bus all the way to Novgorod, watching us.'

The agent, referring to the law-enforcer, is thematic, as it has a major role in the incident described in the passage. The patient, referring to the speaker and friends, is also thematic, as the latter's experience is what the passage is about. Here the agent, besides being thematic, also has some persistence, serving as the subject of the following two clauses.

As mentioned above, there is also a sizable portion of the data, 21.5% (114/531), that is not accounted for in the agentive model. Most of the cases concern passive clauses that are predicted to be active. (26) is an example:

(26) 二少爷在一堆瓦砾中翻蟋蟀翻出了巴掌大一块上好的端砚。砚面上刻有龙飞凤舞的字迹。不知怎么让老先生察觉了，将端砚收走不说，还用竹板奖赏了他的手心。孟吉此刻想那块砚无疑被老先生占有了，用了那样貌似威严却极卑劣的手段。所谓公理是强者为欺负弱者又让被欺者心服而定的。人世间仍是强悍者的天下…

Er   shaoye zai yidui wali   zhong fan   xishuai
second son   at a-pile rubbles in   shuffle cricket

fanchu-le   bazhang da yikuai shanghao   de duanyan.
shuffle-PERF palm   big one-CL top-quality DE ink-stone

Yanmian shang keyou   longfeifengwu de ziji.   Buzhi
ink-top   on   inscribe lively-vigorous DE writing not-know

zenme rang lao xiansheng chajue   le, jiang duanyan   shouzou
how   let old man   discover PRT OBJ   ink-stone take-away

bushuo,   hai yong zhuban   jiangshang-le tade shouxin.
not-to-mention also use   bamboo-board reward-PERF his   palm

Mengji cike xiang <u>nakuai yan wuyi   bei lao xiansheng</u>
Mengji now think that-CL ink no-doubt BEI old man

<u>zhanyou le</u>, yong-le nayang maosi   weiyan   que ji
seize   PRT use-PERF that   look-like dignified but very

beilie   de shouduan. Suowei   gongli shi qiangzhe   wei
despicable DE method   so-called truth is   strong-ones to

qifu ruozhe   you rang ruozhe   xinfu   er   ding de
bully weak-ones and let   weak-ones convinced CONJ set   PRT

'The second son was looking for crickets in the rubbles, and he dug out a palm-size top-quality ink stone. On its top was inscribed lively and vigorous hand writing. Somehow the old man found out. He not only took away the ink, he also hit his hand with a bamboo board. Mengji at this time thought

>that piece of ink no doubt was seized by the old man, using a method that looked dignified but actually was despicable. The so-called truth is set by the strong who bully the weak but also want the weak to be convinced.'

Both the agent, referring to the old man, and the patient, referring to the ink, are thematic in the context, which has to do with how the strong take advantage of the weak. Besides, the agent has a lower RD value than the patient. Therefore, the target clause is predicted to be active, but it comes out as passive.

To summarize, in the active-agentive passive variation, the same four factors that are significant in the active-agentless passive variation are also at work; however, they achieve a significantly lower accuracy rate.

5.   Discussion

Based on the findings above, we can observe a number of characteristics about the variations. First, the use of the *bei* passives is for the most part correctly predicted by adversity and discourse continuity factors, reinforcing the connection between discourse structure and passives, which has been noted cross-linguistically. Agent thematicity is most effective in the active-agentless passive variation, while patient thematicity carries the most weight in the active-agentive passive variation. By contrast, topicality, as measured by RD difference and TP difference, plays a lesser role. TP is not significant, and RD's effect is small. Another aspect of discourse structure tested, namely, local environment, is also not significant. Together, the results show that the choice of the *bei* passive has more to do with cohesion of the broader discourse, rather than local continuity.

Adversity characterizes only 53.7% (295/549) of the passive data. This ratio is compatible with the finding in Xiao et al (2006), where 51.5% of the *bei* passive is considered negative.[8] Nonetheless, it is a strong predictor in both variations, suggesting that it continues to exert strong influence in Modern Chinese, despite the decrease in frequency in the passive. Everything else being equal, an adversative situation is more likely to be expressed in the passive, rather than in the active.

Three issues deserve further comments. I will consider them below: differences between agentive and agentless passives, the role of TP, and the status of the agent and the patient.

5.1   Agentive and agentless passives

One of the major findings is that the variations involving the agentive and agentless passives are significantly different with respect to accuracy of prediction. A source

of this difference comes from thematicity, in particular, the proportion of thematic agents in the two types of passives. This is given in Table 5:

Table 5. Number and percentages of thematic and non-thematic agents in agentive and agentless passives and actives

	agentive		agentless		active	
thematic agent	64	31.8%	36	10.3%	237	71.8%
non-thematic agent	137	68.2%	312	89.7%	93	28.2%
total	201	100 %	348	100 %	330	100 %

Table 5 shows that between the two passives, a much higher percentage of the agent in agentive passives is thematic. The difference is significant (chi-square = 38.09, $p < .0001$). A thematic agent is more like an active agent than a passive agent. Therefore, such an agent is incapable of distinguishing between the active and the passive or predicting the variation. This explains why agent thematicity is less effective in distinguishing the active from the agentive passive.

The findings therefore provide quantitative evidence for difference in usage between the two passives. Even though the two types of passives have been traditionally recognized, to date there has not been much research on whether or how they are used differently. The results also raise an immediate question. What other factors can predict the use of the agentive passive? This will be explored in future work. Semantic factors such as animacy, definiteness, volition, and factors having to do with information structure such as given vs. new, topic vs. focus, are all potential candidates.

5.2 The role of TP

An unexpected result of the quantitative analysis is that in neither variation is TP a significant factor. This seems to go against earlier studies, e.g. Givón (1983), where TP is one of the two major components of topicality, which plays an important role distinguishing actives and passives. How can we explain the discrepancy?

First, however, note that the two types of studies are actually of different nature and scope, and they cannot be straightforwardly compared. As pointed out in Section 1, topicality-based studies of passives are comparative in nature, mainly concerned with how the active and the passive differ in terms of RD and TP. The study conducted here, however, is a variation study and is concerned with predicting when the passive is likely to be used over the active. The types of data collected and the methods used also differ. In this study, the database includes tokens of both forms (e.g. active and passive), and each token has the potential to be expressed in either form. The quantitative analysis carried out is a probability analysis with

a categorical variable, active or passive, as the dependent variable and thematicity, topicality and local environment as predictors. On the other hand, in topicality-based studies, tokens collected need not have the potential to vary between the two forms, the dimensions being considered are aspects of topicality, RD and TP, and the analysis performed is a comparison of mean values. In this context, we can now take a closer look at the role of TP in the two types of studies.

Consider topicality-based studies first. Cross-linguistically, the difference in the mean value between the agent and the patient has been shown to be larger in RD than in TP. This is certainly borne out in Chinese, as demonstrated in Xing's (1993) study. Table 6 is extracted from Xing (1993: 27–28, 30):

Table 6. Average RD and TP values of agents and patients in active and *bei* construction in Xing's (1993) study

	RD	TP	N
active			
agent	3.29	1.38	100
patient	14.15	0.54	100
***bei* construction**			
agent	15.83	0.56	195
patient	5.62	1.49	195

In both active sentences and the *bei* construction, the average RD difference between the patient and the agent (10.96 in active, −10.21 in passive) is much larger than the average TP difference (−0.84 in active, 0.87 in passive). This means that neither in the active nor in the passive is the average TP difference a strong indicator of how the agent differs from the patient. And if we compare the active with the passive, the RD difference is 21.17, while the TP difference is 1.71. This shows that in Xing's study the active-passive distinction is reflected strongly in RD difference, but only weakly in TP difference.

Now we consider the analysis proposed here. The findings presented in Section 4 show that thematicity (agent thematicity in the agentless model, and patient thematicity in the agentive model) carries most of the weight in predicting variations. Both variables correlate somewhat with TP difference (with agent thematicity: 0.42 in agentless model and 0.40 in agentive model; with patient thematicity: 0.153 in agentless model, and 0.12 in agentive model), although none of the correlations are significant. If, however, thematicity is excluded and only topicality is considered, including two variables RD difference and TP difference, the result shows that both are significant factors in the two variations, as given below:

Table 7. Results when only RD and TP are predictors

active — agentless

variable	coefficient	p value
RD difference	−.1576	< .0001
TP difference	.1337	< .0001

active — agentive

variable	coefficient	p value
RD difference	−.1460	< .0001
TP difference	.0802	.0088

Thus it is not that TP does not play a role in the two variations, but that its role is masked by stronger factors, and it only emerges when the latter is not present. Table 7 also shows that the RD difference is slightly more effective than the TP difference. In comparison with Xing's study, we can see that the two studies are not in conflict; they both show that TP has a small role distinguishing the active from the passive.

It is a separate issue why topicality does not compare with thematicity in its ability to predict variations. As far as I know, this phenomenon has not been documented. To better understand the phenomenon, I will take a closer look at how topicality distinguishes the active from the passive in the data. I will only consider the TP difference here, measured in terms of patient TP minus agent TP. The values are of three types: positive (patient has higher TP), 0 (patient and agent have same TP), and negative (agent has higher TP). In active data, negative values are expected, while in passive data positive values are expected (cf. Table 7 above). What is actually observed is given in Table 8:

Table 8. Number of tokens with negative, 0 and positive value on TP difference in three types of clauses

	agentive	agentless	active
negative	36	30	151
0	49	119	79
positive	116	199	100
total	201	348	330

The data does reflect the expected trend, the majority of the tokens being negative in the active and positive in the passive; however, the data also shows two other patterns: First, a portion of the data, 24.1% (128/531) in the active-agentive passive alternation, and 29.2% (198/678) in the active-agentless passive alternation, shows no TP differences; such data is not useful in distinguishing the active from

the passive. Most of the tokens in this group are ones where neither the agent nor the patient has any persistence. In addition, in each of the three groups there is another portion of tokens with values counter to expectation: 30.3% (100/330) of the active, 19.4% (36/201) of the agentive passive and 8.6% (30/348) of the agentless passive. Thus 25.6% (136/531) of the data in the active-agentive passive variation, and 19.2% (181/678) of the data in the active-agentless passive variation go the opposite direction. Such data offsets the effect of data that reflects the general tendency.

Thus in both variations, close to half of the TP data, 49.7% (264/531) in the active-agentive passive variation and 48.3% (328/678) in the active-agentless passive variation, does not support the expected tendency. This may explain why the TP difference is not a strong variable in my data. One contributor to this phenomenon is the fact that in active sentences the patient sometimes persists for quite some distance, while the agent does not. A typical case is when the patient is introduced for the first time in the target sentence, and becomes the topic subsequently, as in the following example:

(27) <u>有一次，齐国国王派遣晏婴去楚国访问</u>。楚国国王听说晏婴富于机智，口才又好，很想找个机会，把他羞辱一番。楚国国王举行一个盛大的宴会，招待晏婴，请了许多文武官员作陪。

<u>You   yici   Qiguo   guowang paiqian Yanying qu</u>
there-is one-time Qi-nation king   send   Yanying to

<u>Chuguo   fangwen</u>. Chuguo   guowang tingshuo Yanying
Chu-nation visit   Chu-nation king   hear-say Yanying

fuyu   jizhi, koucai   you hao, hen xiang zhao ge
good-at wit   eloquence also good very wish find CL

jihui   ba ta xiuru   yifan. Chuguo   guowang juxing
opportunity BA him embarrass one-CL Chu-nation king   hold

yige   shengda de yanhui zhaodai Yanying, qing -le   xuduo
one-CL big   DE party   entertain Yanying invite -PERF many

wenwu   guanyuan zuopei
civilian-military official   accompany

'<u>One time, the king of Qi sent Yanying to Chu to pay a visit</u>. The king of Chu had heard that Yanying was witty and eloquent, and he would like to find an opportunity to embarrass him. The king held a large party to entertain Yanying, and he invited many officials as guests.'

In this example, the agent *Qiguo guowang* 'the king of Qi' has no persistence at all, while the patient *Yanying* occurs a number of times in the following clauses, as it is part of the discourse topic in the following text. Cases like this illustrate that passives are not the only environment for non-prominent agents, the latter can be found in actives as well.

5.3 Patient promotion and agent demotion

The third issue concerns the status of the patient and the agent in the *bei* passives. Previous research has identified two characteristics of the passive cross-linguistically, in comparison with the active: promotion of the patient from object to the subject, and demotion of the agent from subject to object of preposition. This is a view from the syntactic perspective. From a discourse perspective, it would mean that the patient in the passive (as opposed to in the active) is more prominent in discourse, and the agent is less prominent in the passive than it is in the active. This pattern is found in Xing's (1993) study. Is it also found in this study?

The answer is clearly positive for both the agentless passive and agentive passive. Given that thematicity is the most effective factor, I will use it to demonstrate this pattern. Promotion of the patient can be seen by comparing the proportion of thematic patients in three types of clauses, as in Table 9:

Table 9. Number and percentage of thematic and non-thematic patients in three types of clauses

	agentive		active		agentless	
thematic patient	182	90.5%	323	92.8%	228	69.1%
non-thematic patient	19	9.5%	25	7.2%	102	30.9%
total	201	100 %	348	100 %	330	100 %

From the active to the agentive passive, the proportion of thematic patients goes up from 69.1% (228/330) to 90.5% (182/201). The difference in distribution is significant, (chi-square = 31.48, $p < .0001$). From the active to the agentless passive, it goes up from 69.1% (228/330) to 92.8% (323/348). As predicted, the difference in distribution is also significant, (chi-square = 61.08, $p < .0001$). Demotion of the agent can be seen in Table 5, presented in 5.1. The table shows that the proportion of thematic agents plunges from 71.8% (237/330) in the active to 31.8% (64/201) in the agentive passive and 10.3% (36/348) in the agentless passive. Both distribution differences, between the active and the agentive passive, and between the active and the agentless passive, are significant (chi-square = 79.69, $p < .0001$ for the former; chi-square = 263.57, $p < .0001$ for the latter). A comparison between the promotion and the demotion further suggests that the bigger contrast between the

active and the *bei* passive lies in demotion of the agent rather than promotion of the patient, as the former shows a larger gap.

6.   Conclusion

Perhaps the most important discovery of this study is that topic continuity, which has been used as an explanation for the choice of the passive since Givón (1983), turns out to play a minor role in the active-*bei* passive variations. RD has but a small effect, and the effect of TP is not significant. On the other hand, thematicity is the most effective predictor in the variations. This is in line with Thompson's (1987) proposal on when the passive in English is likely to be chosen. The sharp contrast between thematicity and topicality warrants further investigation. Is this pattern also found in other languages? Or is this simply a characteristic of Chinese? These are questions that should be taken up in the future.

The factors considered here do not give a complete picture of the variations, however. In both the active-agentless passive variation and the active-agentive passive variation a portion of data is not yet accounted for. This suggests there are other factors that also contribute to the variations. Identifying possible factors and investigating their significance should be the focus of future studies.

Although adversity no longer dominates the *bei* passive, it is a strong factor in both variations. Interestingly, a new usage of the *bei* passive in recent years seems to be motivated by adversity. In the past few years the usage of *bei* has been expanded. In particular, *bei* has been used to occur with a new set of verbs, including intransitive verbs, such as *jiuye* 'to go to the job market', *hexie* 'to be in harmony', and *xiaokang* 'to be in middle-class'. In the new context the meaning of *bei* actually reverts back to the original negative sense of 'suffer' with a new twist. Thus *jiuye*, which literally means 'to go to the job market', when occurring with *bei*, is interpreted as 'to be reported as having a job when in fact jobless'. So far the new usage appears to be restricted to the agentless passive. On the basis of the findings here, it is predicted that in the new environment the use of the *bei* passive, besides satisfying adversity, will also be conditioned by thematicity. It will be interesting to see whether the prediction is borne out.

Notes

their feedback. I also thank Xia Lin and Jin Zhang for their assistance with the extraction and coding of the data, and John Duchnowski for assistance with statistics.

1. I will refer to the subject of a *bei* passive clause as the patient and the object of *bei* as the agent. They correspond to the object and the subject in a corresponding active clause.

2. These verbs do not carry much semantic content, hence 'light'. For example, *jinxing yanjiu* 'to do research', where the semantic sense is carried by the argument *yanjiu* 'research' rather than the verb *jinxing* 'do'.

3. This does not mean that all telic verbs can occur with *bei*. The examples in (12) simply demonstrate that certain atelic verbs that are incompatible with *bei* can occur in the *bei* passive when they combine with a morpheme signaling a result. Besides *xihuan* 'like', other verbs that take *shang* 'up' include *ai* 'love', *kan* 'look at', and *la* 'pull'. The phenomenon in fact extends to a wide range, although not all, of verb resultative compounds.

4. Examples of verbs of VG include *dangzuo* 'consider as', *shiwei* 'view as', *renwei* 'regard as'.

5. The practice in Givón (1983) is not to count relative clauses as clauses. However, there are no clear justifications for it. I have decided to count them as clauses here.

6. The sequence involving the two idioms, *nianqingqisheng* 'young and energetic', *woxingwosu* 'do one's own way' is treated as a conjunction of two predicates, rather than a conjunction of two clauses. This is because both idioms need to occur with a subject in order to make a complete clause.

7. *mang-zhe ti ta chuli shanhou* 'be busy taking care of aftermath for him/them' is treated as one single clause, which is a serial verb construction. In addition, I take *you* in *haiyou* 'and, also' in this context not as a full verb *you* 'have', but as a grammaticalized marker of existence.

8. Xiao et al. (2006) also include more detailed counts where the data is broken down into 16 genres. The percentage of negative *bei* passive ranges from 77% for news editorials to 19% for official documents.

References

Biber, Douglas, Stig Johansson, Geoffrey Leach, Susan Conrad, & Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.

Birner, Betty J. & Gregory Ward. 1998. *Information status and noncanonical word order in English*. Amsterdam: John Benjamins.

Bresnan, Joan. 2007. "Is syntactic knowledge probabilistic? Experiments with the English dative alternation". In Sam Featherston & Wolfgang Sternefeld, eds., *Roots: Linguistics in search of its evidential base*, Series: Studies in Generative Grammar, 77–96. Berlin: Mouton de Gruyter.

Bresnan, Joan & Marilyn Ford. 2010. "Predicting syntax: Processing dative constructions in American and Austrian varieties of English". *Language* 86.1: 186–213.

Chappell, Hilary. 1986. "Formal and colloquial adversity passives in Standard Chinese". *Linguistics* 24: 1025–1052.

Cooreman, Ann. 1984. "A functional analysis of passives in Chamorro narrative discourse". *Papers in Linguistics* 16: 395–428.

Cooreman, Ann. 1987. *Transitivity and discourse continuity in Chamorro narratives*. Berlin: Mouton de Gruyter.

Givón, Talmy, ed. 1983. *Topic continuity in Discourse: A quantitative cross-language study*. Amsterdam: John Benjamins.

Givón, Talmy, ed. 1994. *Voice and inversion*. Amsterdam: John Benjamins.

Hashimoto, Mantaro J. 1988. "The structure and typology of the Chinese passive construction". In Masayoshi Shibatani, ed., *Passive and voice*, 329–354. Amsterdam: John Benjamins.

Huang, C.-T. James. 1999. "Chinese passives in comparative perspective". *Tsing Hua Journal of Chinese Studies* 29, 423–509.

Li, Charles, N., & Sandra A. Thompson. 1976. "Subject and topic: A new typology of language". In Charles N. Li, ed., *Subject and Topic*, 457–489. New York: Academic Press.

Li, Charles. N., & Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Li, Linding. 1980. "*Bei* zi ju". [The *bei* construction]. *Zhongguo Yuwen* 6: 401–412.

Keenan, Edward. 1985. "Passive in the world's languages". In Tim Shopen, ed., *Language typology and syntactic description*, *Volume 1: Clause Structure*, 243–281, Cambridge: Cambridge University Press.

Myhill, John. 1992. *Typological Discourse Analysis: Quantitative Approaches to the Study of Linguistic Function*. Oxford: Blackwell.

Myhill, John & Xing, Zhiqun. 1994. "A comparison of the function of voice in Biblical Hebrew, Chinese and English". *Language Sciences* 16.2: 253–283.

Rude, Noel. 1988. "Ergative, passive and anti-passive in Nez Perce: A discourse perspective". In, Masayoshi Shibatani, ed., *Passive and voice*, 537–560. Amsterdam: John Benjamins.

Shi, Dingxu. 1997. "Issues on Chinese passive". *Journal of Chinese Linguistics* 25.1:41–69.

Thompson, Sandra. A. 1987. "The passive in English: A discourse perspective". In Robert Channon and Linda Shockey, eds., *In Honor of Ilse Lehiste*, 497–511. Berlin: Mouton De Gruyter

Tang, Sze-Wing. 2001. "A complementation approach to Chinese passives and its consequences". *Linguistics* 39: 257–295.

Ting, Jen. 1998. "Deriving the *bei*-construction in Mandarin Chinese". *Journal of East Asian Linguistics* 7,4:319–54.

Wang, Li. 1945. *Zhongguo yufa lilun*. [Theories of Mandarin Syntax]. Shanghai: Shangwu Yinshuguan.

Wang, Li. 1957. "Hanyu beidongshi de fazhan". [The development of the bei construction]. *Yuyanxue Luncong* 1:1–16.

Wang, Li. 1958. *Hanyu shi gao* [History of Mandarin]. Beijing: Zhonghua Chubanshe.

Wang, Huan. 1957. *Ba zi ju he Bei zi ju* [The *ba* construction and the *bei* construction]. Shanghai: Shanghai Jiaoyu Chubanshe.

Wang, Huan. 1983. "Yingyu he hanyu de beidongju". [The passive construction in English and Mandarin Chinese] *Zhongguo Yuwen* 6, 409–418.

Xiao, Richard, Anthony McEnery, & Yufang Qian. 2006. "Passive constructions in English and Chinese: a corpus-based contrastive study". *Languages in Contrast*, 6.1:109–149.

Xing, Zhiqun. 1993. *Discourse functions of word order in Chinese: A quantitative analysis of diachronic texts*. PhD Dissertation, University of Michigan at Ann Arbor.

Yao, Yao & Feng-hsi Liu. 2010. "A working report on statistically modeling dative variation in Mandarin Chinese". *Proceedings of the 23rd International Conference on Computational Linguistics,* (COLING 2010), 1236–1244.

Zhang, Hongming. 1991. "The grammaticalization of *bei* in Chinese". *Second International Symposium on Chinese Languages and Linguistics* 2:321–360. Taipei: Academic Sinica.

Appendix: abbreviations

ADV	adverb
BA	the *ba* marker in the *ba* construction
BEI	passive marker
CL	classifier
CONJ	conjunction
DE	*de*, possessive, relative, modifier marker
DIR	directional complement marker
DUR	durative marker
EMP	adverb of emphasis
OBJ	object marker
PERF	perfective aspect
PRT	particle
RES	resultative marker

Author's address

Department of East Asian Studies
University of Arizona
Tucson, AZ 85721
USA

fliu@email.arizona.edu