行政院國家科學委員會專題研究計畫　成果報告

中英對比動詞語彙之設計與建造

計畫主持人： 衛友賢

計畫參與人員： 高愷璜、張瑜芬

報告類型： 精簡報告

處理方式： 本計畫可公開查詢

中　華　民　國 92 年 5 月 19 日

行政院國家科學委員會補助專題研究計畫 ■ 成 果 報 告
□期中進度報告

計畫名稱：中英對比動詞語彙之設計與建造

計畫類別：■ 個別型計畫　　□ 整合型計畫
計畫編號：NSC 90 － 2411 － H － 032 － 008 －
執行期間： 90 年 8 月 1 日至 91 年 7 月 31 日

計畫主持人：衛友賢
共同主持人：
計畫參與人員： 高愷璜、張瑜芬

成果報告類型(依經費核定清單規定繳交)：□精簡報告 ■完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢　□涉及專利或其他智慧財產權，□ 一年 □ 二年後可公開查詢

執行單位： 淡 江 大 學

中 華 民 國 92 年 5 月 16 日

On Co-occurrence, Collocation, and Cross-Linguistic Aspects of Lexical Selection in Interlanguage

David Wible[*] and Anne Liu

Graduate Institute of Western Languages and Literature

Tamkang University

[*] Contact author: Please send correspondence to David Wible at dwible@mail.tku.edu.tw.

On Co-occurrence, Collocation, and Cross-Linguistic Aspects of Lexical Selection in Interlanguage

Abstract

An investigation of verb-noun co-occurrence errors in the English production of Chinese-speaking learners is reported. The data reveal two sorts of errors: (1) miscollocations violate conventions of usage which are idiosyncratic to specific lexical items; (2) argument linking violations suggest that learners' misconstruals of verb semantics have regular syntactic consequences. VN miscollocations violate the N's selectiveness toward the V whereas argument linking errors violate the V's selectional restrictions on the N. The latter type motivate a level of representation along the lines of predicate decomposition (Dowty 1979) or event structure (Rappaport Hovav and Levin 1998; inter alia). Both the miscollocations and the argument linking errors pose interesting questions for further research concerning learnability. While these two facets of lexical knowledge are traditionally investigated in relative isolation from each other and under contrasting research paradigms, this investigation is intended to contribute to a model of L2 lexical knowledge that encompasses both.

中文摘要

此項研究成果旨在報告華語學生在英文寫作中所犯的動詞-名詞搭配錯誤。研究資料主要顯示出兩項主要錯誤：（1）搭配詞錯誤違反了各字彙的特定慣例用法；（2）論元連接的錯誤顯示出學習者對於動詞語意的錯誤解釋是受到句法上的影響，且此影響是具有規則性的。動詞-名詞搭配錯誤違反了使用名詞時的動詞選擇，而論元連接錯誤則是忽略了在使用動詞時，可搭配名詞的選擇限制。後者並證實了述詞分解（Dowty 1979）或事件結構（Rappaport Hovav and Levin 1998; inter alia）等類似呈現層次在語法上存在的必要性。搭配詞錯誤與論元連接錯誤兩者都提供了未來關於可學習性（learnability）方面的研究許多有趣的問題來探討。雖然在過去的研究裡，這兩項字彙知識都被分隔開而獨立研究於相對的研究領域中，此項研究的目的即是提出一個可以涵蓋這兩項第二語言字彙知識的研究模式。

1. INTRODUCTION

One persistently difficult aspect of lexical knowledge for second language learners to master is co-occurrence restrictions on the distribution of lexical items. In this paper, we focus on a tightly circumscribed set of data as a window onto this particular aspect of lexical knowledge in the interlanguage of L2 learners. We look specifically at errors concerning English verb-noun (VN) co-occurrences. In our analysis of learner corpus data we end up claiming that one class of these lexical errors are best seen in terms of collocations and another class in terms of argument linking regularities arising from the semantics of the verb. These two facets of lexical knowledge are traditionally investigated in relative isolation from each other and under contrasting research paradigms. We hope that our exploration of the two within a narrow range of data will help illustrate that both will play a role in an optimum model of L2 lexical knowledge.

The paper is organized as follows. In section 2 we describe our data and its sources. Section 3 focuses on collocation and presents certain patterns that emerge concerning the syntagmatic relationship between the N and the V in the VN errors that we analyze. Moreover, we try to articulate there certain overlooked challenges that collocation knowledge poses for the issue of learnability. In section 4 we consider a subset of the learner VN co-occurrence violations, turning from collocation to the notion of predicate-argument relations and argument linking. Our proposal, consistent with Juffs (1996) and Montrul (2000; 2001), suggests that learners' misconstruals of the behavior of verbs in L2 are sensitive to particular components of the semantics of those verbs, components which can be made transparent through predicate decomposition along the lines of Dowty (1979) or, consistent with Dowty's approach, possibly by exploiting event structure along the lines developed by Rappaport Hovav and Levin (2001) inter alia. A discussion of the implications of the data with respect to the ontology of interlanguage grammars is presented in section 5. Section 6 concludes the paper.

2. THE DATA

The data that we analyze consists of 260 non-target English VN co-occurrence tokens produced in the writing of English learners in Taiwan whose primary language is Chinese.   These tokens were extracted from a learner corpus of English called *English Taiwan Learner Corpus (*or *EnglishTLC)* currently comprising over one million words (tokens) of written English by high school and university students in Taiwan. *EnglishTLC* is integrated into an online English language-learning

platform known as IWiLL (see Wible et. al. (2001) for details on the design and functionality of various modules of IWiLL). The architecture of the platform has been designed to facilitate the sorts of research that we are reporting here. The miscollocations analyzed were all extracted from the learner corpus by means of a bootstrapping heuristic developed by the research team that created the IWiLL platform.[1]

Using this approach, we automatically extracted 1,100 tokens of student errors marked online by teachers as lexical errors. An examination of these tokens by hand uncovered 292 tokens of lexical co-occurrence errors were found among them. These were then categorized according to the part of speech (POS) of the two co-occurring lexical items. The results are shown in Table 1.

Miscollocation type	Frequency
V N	260
Adj N	25
V Adv	5
Adv Adj	2
Total	292

Table 1: Miscollocations by POS type and frequency

Note that nearly 90% of the these error tokens marked by teachers (260 out of 292) are VN pairs. This striking predominance of VN co-occurrence errors is what initially led us to focus on them for this study. In the following section, we discuss some facts that came to light upon an initial analysis.

3. THE SYNTAGMATICS OF VN MISCOLLOCATIONS

In this section, we analyze the miscollocations by considering the syntagmatic relationship between the N and V of each pair, first to determine the grammatical function played by the N in relation to the V, (subject, direct object, etc.) and then to determine which member of the VN pair is the miscollocate (i.e., the wrong word). In both of these analyses, the data shows overwhelming tendencies clearly suggesting patterns that call for explanation.

3.1 The Miscollocations and Grammatical Function

First we analyze the NV miscollocations for the functional relations that stand

between the N and V in each pair. Specifically, we determined the grammatical function of the (NP headed by the) noun in relation to the verb in each NV miscollocation pair. The grammatical relations we considered were: subject, direct object, indirect object, oblique (i.e., adjunct). The results, shown in Table 2, are striking. The noun in the VN pair is an argument of the verb in 254 out of the 260 NV miscollocations (i.e., in more than 97% of them), and the noun is a non-argument (adjunct) of the verb in only six of the 260 NV miscollocations (i.e., in less than 3% of them). Considering the 254 cases where the noun heads an argument of the verb, in 242 out of those 254, the noun heads the direct object, in ten cases it heads the subject, and in two, the indirect object. Moreover, all ten cases where the miscollocate noun heads the subject are passive constructions; in other words, even in these ten cases, the noun heads the deep object (which has been raised to subject). Put another way, in 252 (242+10) out of 260 cases, the noun in the VN miscollocation heads the direct internal argument of the verb of that VN pair. In the other eight cases, it heads an indirect object (two cases) or oblique (six cases). There are no cases where the N heads the deep subject of the V.

Argument Status	Frequency
The N heads an argument of the V	254
direct object	*242*
subject	*10*
indirect object	*2*
The N is part of an adjunct	6

Table 2: The distribution of the grammatical relations in VN miscollocations.

There are two possible reasons for absence of subject miscollocates and the overwhelming predominance of direct internal arguments of the verb as miscollocates. First, it could be a phenomenon of acquisition. That is, it could be that while English has collocations which involve verbs in collocational relations with their grammatical subjects, learners simply have little difficulty with these compared to verb-object collocations.

An alternative view of the absence of subject-verb miscollocations in the data would treat it as a reflection of the target language rather than of acquisition. That is, on this view there is simply a dearth of subject-verb collocations in English and this would account for the dearth of violations of subject-verb collocations by learners. On such an account, independent of learner miscollocations, the NV collocations of English (or universally) are predominantly collocations between the V and its internal direct argument (its deep object) and verbs simply rarely enter into collocation

relations with their underlying subject arguments. With few or no subject-verb collocations in English, there would be correspondingly few on no violations of them by learners.

While a full-blown investigation comparing these two views and the relevant data would take us beyond the scope of this paper, we conducted an informal survey of an English collocation dictionary which is highly suggestive. By sampling the *Longman Dictionary of English Collocations* (LDEC), we found that Ns occurring in English VN collocations indeed are predominantly direct objects and rarely subjects. Specifically, we examined 100 nouns that participate in VN collocations in LDEC and found that 80 of the nouns serve as direct object in (one or more) collocations with a verb whereas only 12 of those 100 nouns serve as subject in any collocations with verbs.[2] Many of these nouns participate in more than one collocation (e.g., *take medicine*; *prescribe medicine*); it turns out that the 100 nouns examined participate in a total of 369 NV collocations. We found that the noun serves as direct object of the V in 80.4% of those 369 collocations (or in 297 of them) but as subject in only 8% (or 30) of them.

These figures would suggest then that it is plausible that learners' miscollocations are skewed toward NV pairs in which the noun serves as a grammatical object rather than subject of the verb simply because the proportion of English NV collocations are skewed that way as well rather than because learners have more difficulty with acquiring object-verb collocations per se compared to subject-verb collocations.

3.2 The Miscollocations and Part-of-Speech

Here we examine the NV miscollocations to determine which of the two lexemes in the NV pair was the miscollocate and which one the focal word.[3] The null hypothesis would be that we should find an even split between noun and verb as miscollocate; that is, it would be the noun which is the miscollocate in about one half of the 260 miscollocations, and it would be the verb which is the miscollocate in the other half of the cases. It turns out, however, that in over 97% of the 260 NV miscollocations (253 of them), it is the verb which in the miscollocate, and in less then 3% (only seven cases) is it the noun which is incorrect. What we are seeing here is that nouns (overwhelmingly grammatical objects) are picky about the verbs they can appear with. The miscollocations here arise when learners are not sensitive to this choosiness; they produce verb-object miscollocations due to choosing the wrong verb for that noun. Interestingly, this appears to be the converse of what we saw above in looking at the grammatical functions of the N in the NV miscollocations. There we were faced with the fact that overwhelmingly the N was the (head of the)

grammatical object of the collocate verb and in no case was it the (head of the) underlying (semantic) subject of the verb. That phenomenon appears closely tied to the verb's selection of its arguments, specifically, that verbs directly select their objects but not their subjects. This, in turn, has to do with the verb's selectiveness toward its arguments. By looking at the same set of VN miscollocations to see which element—the N or the V—is the miscollocate, however, we find the converse of this direction of selectiveness, that is, we see the argument noun's selectiveness toward the verb it appears with. Specifically, in the VN miscollocations in our data in an overwhelming proportion of the cases (97%), it is the verb, not the noun, which is the miscollocate.

This latter choosiness of nouns toward the verbs that select them fits in with the picture developed by a number of linguists in the semantics literature which shows a mutually constraining relation between a verb and particular arguments that it "selects." Bach 1986, Partee 1987; Filip 1994, Hinrich 1985, Krifka 1992 inter alia deal specifically with aspectual facets of meaning such as telicity and how properties of the internal argument can determine or coerce the aspectual class of the verb that selects it. More directly relevant to our findings is the work of Pustejovky (1995) on a model for a generative lexicon. He shows how a range of surprisingly subtle semantic and morphological properties of nouns govern the interpretation of the verbs that select them. For example, the same verb *enjoy* denotes clearly different activities in each of the following sentences.

> (1) I enjoyed the book.
> (2) I enjoyed the ice cream.

Here it is obviously the choice of object which determines that in (1) *enjoy* entails the act of reading but in (2) is entails eating instead.

Pustejovsky is concerned with the semantic effect which an argument noun can wield over a polysemous verb, restricting the interpretation of that verb. While he shows the role of nominal arguments in coercing particular senses of a single verb, our learner miscollocation data suggests something parallel but different: the influence of the nominal arguments in selecting among different verbs rather than among different senses of the same verb.

3.3 Collocation, Input, and Learnability

Before turning from miscollocations to the data related to argument linking, it is worth pointing out the relevance of miscollocation to some issues of learnability in both L1 and L2 acquisition. Collocation is rarely if ever included in discussions of learnability theory. Perhaps this is because it seems obvious that since collocations

are idiosyncratic conventional lexical pairings (or groupings), the issue of acquisition then is trivial: collocations must be acquired by brute force of exposure to co-occurrences of the collocating lexemes in linguistic input. It is far from obvious, however, what sort of input could constitute evidence for a collocation and lead a learner, therefore, to expunge a miscollocation from their interlanguage. Consider *eat medicine*. What sort of input would lead learners to drop this miscollocation from their repertoire? Would it be the presence of *take medicine* in their input? If so, then what is to prevent the same from happening in parallel cases, for example, what would prevent the presence of *purchase medicine* in the input from leading the learner to assume incorrectly that the paraphrase *buy medicine* is a miscollocation? Would the learner need to wait until subsequently encountering *buy medicine* to conclude it is not a miscollocation? Moreover, in many cases VN collocability seems to be inherited by the hyponyms of the N despite the dearth of relevant VN tokens in the input. The verb *take*, for example, collocates not only with the noun *medicine*, but with *penicillin*, *aspirin*, *antihistamines*, *antibiotics*, and so on down to the proper names of specific drugs. It is highly doubtful that the collocability of each of these with *take* is reducible to the sheer statistical aspects of the co-occurrences of each of these nouns with *take*. It is correspondingly implausible that language users need to wait for co-occurrences of the verb *take* with each particular noun naming a medication before the learner would know that that noun collocates with *take*. In other words, language users' collocation knowledge almost certainly exceeds what is directly attested in their linguistic input.

Thus, while the acquisition of collocations is certainly heavily dependent on L2 input, it is far from clear how. The presence of collocations in the input can indicate to the learner that a particular co-occurrence pairing is acceptable, but that is different from indicating to the learner that the pairing is, more specifically, a collocation and that at least one of the words in the pair is not interchangeable with its near synonyms (*spend time* vs *pay time*). While statistical research on detecting collocations relies on the syntagmatic dimension by detecting statistically significant co-occurrences with measures such as mutual information (MI) (see Krenn 2000; Kilgraff and Tugwell 2001; Manning and Schutze 1999; Shimohata 1997; inter alia), early work on collocation acknowledges the paradigmatic dimension as well. As Halliday has pointed out, we sense *strong tea* as a collocation not simply because of this syntagmatic co-occurrence, but because we reject paradigmatic alternatives such a *powerful tea* (1966:150-52). What precise elements in the linguistic input, then, can constitute evidence of a collocation and distinguish collocations from mere co-occurrences for a learner? This is an important question for both first and second language acquisition which to our knowledge remains largely unarticulated and

unexplored.

4. ARGUMENT SELECTION AND THE SYNTAX-LEXICAL SEMANTICS INTERFACE

In the previous section, we looked at the miscollocation data in terms of syntagmatic distributions, focusing on the relation of the N to the V within the VN collocation pairs.   While what is typically emphasized with collocations is their syntagmatic dimension--the relation holding between the co-occurring or collocatable elements--second language data casts their paradigmatic dimension into relief as well. A substantial number of miscollocations have one or more corresponding terms that would be acceptable replacements for the miscollocate. We consider here the paradigmatic relationship that stands between the miscollocate V produced by learners and the counterpart correct collocate V that would replace it.   For example, in the miscollocation "pay time," where the correct expression would be "spend time," we consider the relationship between the miscollocate ("pay") and the counterpart correct collocate ("spend") to see if there are any generalizations to be captured. In other words, in this section we consider the vertical relationship (between V and V) represented by the vertical arrow in Figure 1.
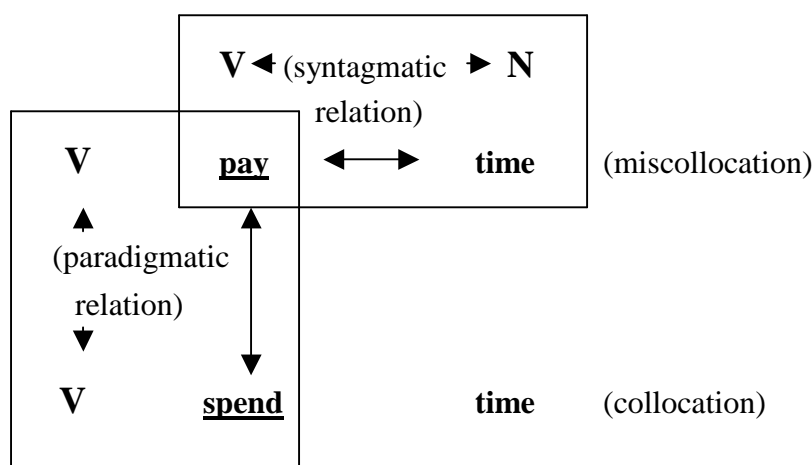

Figure 1: Syntagmatic and paradigmatic relations among *pay*, *spend* and *time*

Liu (2002) investigates in detail the entire set of 260 NV miscollocations tokens described in section 3 above. In this paper, we focus on a subset of these errors that show a pattern concerning the syntax and lexical semantics of the misused V and its correct V counterpart. We make use of some analytical tools from lexical semantics, mostly predicate decomposition, and the phenomenon of verb diathesis.

The VN co-occurrence errors among those extracted from the learner corpus

which we focus on are shown in Table 3.   First, we should point out that this class of VN pairs is a narrow subset of all the 260 VN errors that we extracted from the learner corpus. On the one hand, they are syntactically unique among the entire set of 260 VN error pairs in that these contain the only verbs which differ from their corresponding correct verb with respect to transitivity.   An example is the learner sentence *I have <u>remained</u> the hairstyle for three years*. The intransitive verb *remain* is used transitively by the learner with a direct object (*the hairstyle*). The acceptable counterpart verb corresponding to this miscollocate intransitive *remain* is a transitive verb, *keep* or *preserve* or *maintain*. In this respect, as we narrow our focus to this subset of the data, these VN pairs seem to be less about collocation and more about lexical selection and argument linking than the VN pairs we considered in section 3 above. This difference should become clear as we proceed, but the shift is worth pointing out here.

On the other hand, we would like to point out in what follows that this set of sentences involving misuse of transitivity exhibit two other properties--one syntactic and one semantic--which set them apart from the other 260 VN errors in our data. First, the syntactic property is that the misused verb differs from its correct counterpart verb(s) not simply with respect to transitivity, but also with respect to argument linking. None of these errors in transitivity simply entail the omission or addition of an object. Rather, as in the learner sentence *I remained the hairstyle for three years*, the grammatical object in the learner's use of *remain* corresponds to the grammatical subject in the target use of that verb: *<u>The hairstyle</u> remained*. In other words, by comparing the learner's use of the verb to target use of the verb, we see not simply a transitivity alternation, but rather a sort of ergative alternation (pictured in Figure 2). And this pseudo ergative alternation holds for each of the verbs that show the learner misconstruing its transitivity.

V1^{trans} ⇔ (cause (V2^{intrans}))		Freq of miscoll	Example learner sentences
show/ reveal	← (cause (*appear))	9	■ Her cute puppy started barking loudly to *appear* his fear. ■ Hiding just *appeared* their insignificancy.
keep/ preserve leave	← (cause (*remain)) (*proximal*) ← (cause (*remain)) (*distal*)	8	■ I have remained the hairstyle for three years. ■ Although Margalo left after days, she still remained a vivid image in Stuart's mind and became his motivity to get away from his home.
raise arouse	← (cause (*rise/*arise)) ←(cause (*rise/*arise))	3	■ He will design such activities as role-play, c classroom-debate, or speech contest to *rise* students' interest. ■ The commercial takes advantage of entertainers' popularity to *arise* people's interest to buy the product.
preserve	← (cause (*last))	1	■ Nevertheless, selfish people hard to *last* this beautiful island forever.
* cause	→ (cause (occur))	1	■ Lots of aftershocks *caused* constantly for several days.
stir	← (cause(*arise))	1	■ So, the main action will let readers try to think and will *arise* their imagination.
achieve	← (cause (*come true))	1	■ I will work hard to *come true* dreams.
* shed	→ (cause (fall))	3	■ the tears couldn't help shedding down.

Table 3. Causative Relation between Miscollocate V Target V
(Miscollocate V is indicated by *)

```
Learner transitive use of V1:        *She   remained   [the
hairstyle]



Target intransitive use of V1:    [The hairstyle] remained.
```

Figure 2: The ergative-like relation holding between the learner's transitive use of *remain* and target intransitive use of *remain*

Secondly, in addition to this regular syntactic correspondence, all of the miscollocation cases that exhibit this correspondence concerning transitivity and argument linking also share a common semantic correspondence to their counterpart correct verbs. Specifically, these miscollocate verbs and their counterpart correct collocate verbs are causative/non-causative counterparts of each other.[4] This convergence of a syntactic correspondence with a lexical-semantic one leads us to the interface of syntax and lexical semantics in the representation of interlanguage.

Before considering the data in further detail, we sketch here a relevant hypothesis concerning interlanguage and the syntax-lexical semantics interface. Research on interface between syntax and lexical semantics is primarily concerned with finding regular correspondences between the meaning of predicates on the one hand and the syntactic expression of their arguments on the other (See Perlmutter 1978; Grimshaw 1987 and 1990; Levin 1985 and 1995, Rappaport et al 1993; Pinker 1994; inter alia). The assumption is that the syntax of a verb can be predicted to a great extent from its semantics (Pesetsky 1982; Wasow 1985; Hale and Keyser 1993; Lasnik 2000:132-33). Turning to second language acquisition, this assumption potentially leads to certain predictions. Specifically, if particular aspects of the syntactic behavior of verbs derive from their semantics and, crucially, if second language learners come to the acquisition task sensitive to this correspondence, we would expect that if a learner misconstrues relevant aspects of a verb's meaning in the target language, this should carry syntactic consequences in their use of that verb.[5] In the data we consider below, we seem to find support for just such a view.

What we are pointing out with respect to the data in Table 3 is that these verb pairs of wrong verb and correct verb counterpart share a common pattern deeper than simply a difference in transitivity. That is, the syntactic correspondence whereby the incorrect verb is intransitive and its correct counterpart verb is transitive correlates with a semantic correspondence between the two with respect to causativity. Specifically, whereas the meaning of the incorrect verb is non-causative in the target language, it is entailed in the meaning of the corresponding acceptable verb, which is causative. Conversely, in the two cases where the meaning of the wrong verb is causative in the target lexicon, it entails the meaning of the counterpart acceptable verb(s). To see the semantic correspondence, consider again *I have remained the hairstyle for three years*. The incorrect verb *remain* is both intransitive and non-causative. An alternative verb which would be an acceptable substitute in this same syntactic frame—*kept, retain, maintained*—would be transitive and causative and, moreover, entails the meaning of *remain*. That is, the correct causative

transitive [*to keep X*] can be decomposed as [*CAUSE* [*X remain*]], where the decomposition transparently entails the corresponding intransitive, non-causative incorrect verb *remain.*

Similar non-target transitive/causatives have been noted in earlier literature. Rutherford (1987: 88-89) cites learner examples such as *This construction will progress my country* and *The shortage of fuels occurred the need for economical engine.* Rutherford points out that the learners are introducing here both a transitive and causative use that is non-target, though he does not mention the ergative-like alternation between the non-target and target use of the verbs. Yip (1994:129 and 1995:137-38) cites Rutherford's examples and notes in addition the ergative connection (i.e., that the subject in the target intransitive corresponds to the object in the learner's non-target transitive counterpart). She cites these in the context of trying to account for learners' overuse of passive without noting any connection between causative meaning to transitive syntax in the learners' grammar.

Montrul (1997; 1999; 2000) and Juffs (1996) investigate L2 learners' transitivity errors. (Only Juffs considers Chinese learners of English.) Both researchers look at the role of transitivity alternations (verbs which permit both an intransitive-non-causative use and a transitive-causative counterpart), investigating whether learners' non-target transitivization of intransitive verbs is the result of either transfer from alternating verbs in their L1 or overgeneralization from alternating L2 verbs to non-alternating L2 verbs. We will propose that a simpler hypothesis might account for at least the sorts of transitivity errors in our data. Specifically, our proposal requires no reference alternating verbs but only to the generalization that causative verbs are transitive. That is, we bypass any comparison with verbs that allow transitivity alternations in accounting for learners' transitivity errors.

Oshita (1997) cites the Rutherford (1987) data and suggests that the learners' addition of causative meaning to a non-causative intransitive verb entails an alternation in the verb's predicate argument structure (PAS). As we note below, however, PAS captures only the number and ordering of arguments without reference to the semantic role that causativity plays in governing the learners' alternation. That is, an optimal account will capture the correspondence between causative semantics and argument linking exhibited in the learner data.

To capture this correspondence we suggest representing the lexical semantics of the verbs through predicate decomposition along the lines used in Dowty (1979) and pursued by a number of other linguists in the form of event structures. There are clear advantages to such a representation. First, a level of representation for predicate decomposition is independently motivated, and so we are not introducing it simply for insights into this class of learner miscollocations. With respect to the interaction of

syntax and lexical semantics, Dowty (1979), Rappaport Hovav and Levin (1998); Rappaport, Laughren, and Levin (1993), Jackendoff (1990) Pinker (1989) and others provide a range of evidence motivating a level of lexical decomposition. Specifically, it is shown that such a level of representation makes it possible to capture generalizations concerning the semantics of verbs on the one hand and the mapping of their arguments onto the grammatical relations expressed in a clause on the other.[6]

Second, without such a level there would be no way to capture the relationship between the incorrect and correct co-occurring verb shown in Table 3. That is, we would need to treat these pairings as fortuitous, each pairing either arbitrary or at least distinct in type from all of the others. With predicate decomposition, however, we have a simple and independently motivated means of capturing a generalization about how the verbs in each verb pair in Table 4 are related to each other and the nature of the learners' misconstrual. Specifically, in the example where *remain* is the miscollocate and *keep* is one of the acceptable counterparts, we can capture their correspondence in the target language system informally as follows:

[X keep Y] => [X CAUSE [Y remain]]

In other words, *keep* is represented as a causative transitive verb and *remain* as a non-causative intransitive counterpart entailed in the representation of the causative. This is consistent with observation of Rappaport Hovav and Levin (2001) that "(a)s sentences with a transitive use of a lexical causative verb V are fairly well paraphrased as 'cause to V-intransitive', such verbs are usually assigned an event structure involving two causally related sub-events (e.g., Dowty 1979:91-94, Levin & Rappaport Hovav 1995:83, Parsons 1990:109-11)" (p.783). What we are suggesting is that positing such a complex event structure for learners' transitive use of strictly intransitive verbs makes it possible to capture generalizations concerning these learners' interlanguage.

How does this contribute to the analysis of the misused verbs under consideration? We can informally illustrate its relevance as follows. Without decomposition, we are comparing the predicates 'keep' and 'remain' under (a) and (b) in (3) below, whereas with predicate decomposition, we are comparing the same verbs under (a) and (b) in (4), where the relationship becomes transparent and the generalization can be stated.

(3)   a. [Y remain]   (4)   a.   [Y remain]

      b. [X keep Y]   b.   X CAUSE [Y remain]
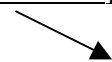
Notice that this causative-non-causative correspondence holds for each of the cases in Table3. We represent those verb pairs in Table 4. The pattern here is that the

learner is using either the intransitive non-causative predicate in place of the transitive causative counterpart whose meaning entails the intransitive predicate that is being misused, or, in the case of 'shed' and 'cause' precisely the reverse: using the transitive causative verb rather than the appropriate intransitive non-causative. The correspondence captured here regardless of the direction of the error suggests that in all of these cases there is a regularity behind the learners' production.

Keep – *Remain (hairstyle)	[X keep Y] = [X CAUSE (Y remain)]
Arouse - *Rise (students' interest) Arouse - *Arise (people's interest)	[X arouse Y] = [X CAUSE (Y rise)] [X arouse Y] = [X CAUSE (Y arise)]
achieve - *come true (dreams)	[X achieve Y] = [X CAUSE (Y come true)]
Occur - * (aftershocks) cause	[X cause Y] = [X CAUSE (Y occur)]
Preserve – *Last (this island)	[X preserve Y] = [X CAUSE (Y last)]
Drop – *(tears) shed	[X drop Y] = [X CAUSE (Y shed)]
Stir – *Arise (their imagination)	[X stir Y] = [X CAUSE (Y arise)]
Show – *Appear (his fear)	[X show Y] = [X CAUSE (Y appear)]

Table 4. Miscollocate/collocate verb pairs and their semantic entailment relation

We can venture a stronger claim based upon the widely noted fact that lexical causative verbs are transitive in English. This suggests that the regularity could derive from the semantics of the predicates and the syntactic realization of their arguments. That is, when a learner misconstrues the semantics of the verb, taking a non-causative intransitive verb for a causative, the syntax of the error (i.e., using the verb transitively) would follow. This rests on the assumption, however, that learners themselves take all lexical causatives to be transitive.[7]   Such an assumption would be highly plausible if it were not a language specific fact, but a universal: Lexical causative verbs are transitive universally. This is in fact a common assumption in the literature on causatives (Fodor 1970; Pinker 1989; Shibatani 1976; inter alia). It would be, then, plausible as a default assumption in the hypothesis space that the learner brings to the L2 acquisition task.[8]

On this view, our data and the analysis that we propose for it are relevant to the theory of the syntax-lexical semantics interface. In this respect, our proposal is similar in spirit to the work of Montrul (1997; inter alia) and Juffs (1996). As we mentioned above, research on this interface has been motivated to a great extent by the hypothesis that the syntax of predicates is predictable from particular aspects of

their semantics. We would like to note three points in this regard.

First, such a hypothesis can derive predictions for interlanguage. That is, if L1 learners can derive the syntax of the verbs they are learning from elements of the meaning of those verbs, then it should be the case for L2 learners as well. Moreover, it should be the case that wherever learners misconstrue the meaning of a predicate in respects that impinge on its syntax, in those cases we should see corresponding misuses of the syntax of that predicate. It has only been recently that such implications of event structure or lexical decomposition for L2 acquisition have been pursued in any detail (Juffs 1996; Montrul 1997; 2000).

The second point is that this prediction appears to be borne out by the transitivity errors in our data. Wherever a learner construes a non-causative English verb as causative, that learner also shows a corresponding transitive use of that verb. Hence, the striking correspondence that we find between semantic misconstruals of a verb with respect to causativity on the one hand and syntactic misuses of the verb with respect to transitivity and argument linking on the other is to be expected under the view of the syntax-lexical semantics interface which derives verb syntax from verb semantics.

Third, the learner data we are considering suggest that parceling the semantics of a verb into certain primitives at some level of representation that is available to the syntax may account for certain interlanguage phenomena. Levin and Rappaport (1998) suggest that such decompositions and the notion that verb meanings are componentially "built" from such pieces monotonically raises the possibility of predictable types of cross-linguistic variation, for example, that semantic components which are lexicalized as part of verb roots in one language could in another language be expressed as distinct morphology (See also Talmy 1980). A parallel implication for interlanguage is that predicate decomposition raises the possibility that learners' lexical acquisition could be shaped by whether or not the causative component of a causative verb has been acquired or, conversely, whether a causative component has been posited by a learner for a non-causative verb.[9]

We are suggesting then the possibility of a regular syntax-lexical semantics interaction at work in interlanguage, specifically, that the L2 learners' lexical semantic construal of a verb has coattails which carry syntactic consequences. There have been a variety of proposals concerning levels of representation where syntax and lexical semantics intersect. Here we briefly distinguish two sorts of proposed representation types and show their relevance to our analysis. One sort of representation is sometimes referred to as Predicate-Argument Structure (PAS) (Grimshaw 1990, inter alia). The other type would include Jackendoff's Lexical Conceptual Structure (LCS) and related work on Event Structure (Rappaport Hovav

and Levin 2001 inter alia.). What all of these have in common is some means of representing semantic aspects of lexical items (usually predicates and most frequently verbs) which impinge upon the syntax of the clauses those lexemes appear in, specifically with respect to argument linking. What distinguishes the two sorts of representation is the type of information and level of detail they encode. This difference affects our choice concerning which is a more appropriate representation for the lexical semantic aspects of interlanguage that our data is pointing toward, and so it warrants some comment here.

PAS is essentially a sort of predicate calculus inspired by Frege which encodes the number of arguments required by the verb and gives these arguments a structured representation so that they can be mapped onto the grammatical functions of a clause in some predictable way. In addition, a separate Thematic Hierarchy is usually required to determine how semantic roles map onto the arguments encoded in the PAS. Essentially, this family of representations restricts itself to encoding information about the number, type, and ordering of arguments of the lexical head.

LCS or event structure representations, on the other hand, are more elaborated than PAS or theta grids and are motivated by the assumption that in addition to the number, type, and ordering of arguments, there are other aspects of a verb's meaning which are grammatically relevant and are shared by entire classes of verbs. Accordingly, Rappaport Hovav and Levin (1998) describe event structure as the "grammatically relevant lexical semantic representation" (p. 106). Dowty's predicate decomposition (1979) can be seen as belonging to this family of approaches. Common to these event structure approaches is a set of primitive predicates in the representation vocabulary such as CAUSE, BECOME, DO (or ACT), and semantic constants such as STATE, LOCATION, and some calculus or template for combining them with arguments.

Two motivations for positing a level of event structure are that (1) there are highly restricted components of meaning beyond argument structure which are shared by classes of verbs and have predictable consequences for their syntax (Rappaport Hovav and Levin, 1998), and (2) learners need access to the sort of information encoded in the representations of these components in order to acquire the syntax of these verbs (Pinker 1989).

The reason for the above excursion into the theory of lexical semantics representations is to fit our analysis into this broader picture. In order to capture the generalizations we seem to detect in the learner data, we need an interlanguage lexical semantic representation richer than those of the PAS-type alone. That is, we need something along the lines of proposed event structure representations. This is because the generalizations we are trying to capture must refer to semantic primitive

predicates unavailable in PAS, at least to the semantic notion of CAUSE.[10]

5. OVERGENERALIZATION AND TRANSFER ACCOUNTS

Notice that the analysis in Section 4 is stated in terms of the target language with no reference to the learners' L1. Taking L1 into account would seem to leave open two basic types of analysis for this sort of error: overgeneralization or transfer. We sketch here the general shape of the two most obvious approaches and certain reasons to believe that they are inadequate, that is, that the optimal account, whatever it turns out to be, will need a more enriched set of interlanguage concepts beyond the straightforward notions of transfer and overgeneralization.

5.1 L1 and a Transfer Account

One straightforward account would be to attribute the errors in Table 4 to transfer to L1 influence. This could be the case if the Chinese equivalents of the misused verbs in Table 4 permitted a transitive use. There is a striking property of Chinese which we have come across in looking into the relevant Chinese data concerning this possibility, however, which would seem to undermine a transfer analysis. It appears from our initial exploration that Chinese has no morphologically simple verbs which are causative transitives.[11] Such causative meanings must be expressed periphrastically or as a compound, with one verb expressing the action and another predicate denoting the "caused" result. For example, the Chinese predicate for *break* is 'po'. But it can only express the inchoative intransitive meaning of break: 'Bei-ze po-le' *The glass(es) broke*. The causative transitive requires an additional verb expressing the action, and 'po' becomes the resultative complement to that predicate denoting the resulting state. *Ta ba bei-ze da-po-le* "S/he broke the glass(es)." Or literally *S/he hit the glass broke*. Put another way, Chinese has no simple diathesis pairs like the English causative/inchoative alternating verbs such as *break, sink*, etc. Any such alternations in Chinese involve the introduction of corresponding morphological changes as well. Our point here is this cross-linguistic difference casts doubt on a transfer account of the transitive-causative learner data in Table 4. There is no straightforward analogy in L1 for what the learners have produced in L2 with the verbs in Table 4.

5.2 L2 and Overgeneralization Accounts

There is more than one possible over-generalization approach to the data in Table 4. Here we consider two basic sorts. On one overgeneralization approach to the data,

the learners are overgeneralizing a verb alternation. English has a well-known class of verbs which exhibit a causative-non-causative diathesis that entails a transitive-intransitive syntactic alternation. Verbs such as *break*, *sink*, *drop*, *melt*, and many others have both a transitive causative and an intransitive non-causative use, as show in (1-2)

      (5).   a. <u>The toy</u> broke. →   b. He broke <u>the toy</u>.

      (6).   a. <u>The ship</u> sank. →   b. They sank <u>the ship</u>.

In each alternation in (5-6), the underlined argument is shared by the (a) and the (b) version: the object of the transitive corresponds to the subject of the intransitive version; hence these are sometimes referred to as *ergative* alternations. One type of overgeneralization account of the errors in Table 4 would posit that learners overgeneralize this alternation to verbs which do not alternate in the target language. That is, the learners realize that English has verbs which permit an ergative alternation and incorrectly overgeneralize this alternation to the verbs in Table 4. For example, on analogy with the alternating ergative verb *break,* the learner treats the non-alternating intransitive verb *remain* as having a transitive causative counterpart as well, leading to *I remained the hairstyle* by overgeneralization.

    Again, on this straightforward overgeneralization view of the data in Table 4, what is being overgeneralized is a verb alternation. Such an analysis, however, risks running afoul of the 'comparative fallacy' (Bley-Vroman 1983) because the analysis can capture the generalization in the learners' relevant output only when framed in terms of the target language system. In other words, the account of the learner's error is plausible only assuming the learner knows that the misused verbs such as *remain* and *appear*, for example, are indeed intransitive in English. If they do not know that the verb which they are misusing transitively has an intransitive use, then, seen in terms of their own interlanguage, there would be no transitive/intransitive "alternation" involved. We have shown no evidence, however, that each learner who produces one of these unacceptable causative transitives realizes that the verb in question has an acceptable intransitive non-causative use.[12]

    There is another sort of overgeneralization account that would make no claims about whether the learner realizes that the misused verb is intransitive in English and would seem to circumvent the comparative fallacy. Rather than overgeneralizing a verb alternation (NP1 V NP2 → NP2 V), the learners could be using the verb transitively without regard to whether it has an intransitive counterpart. They could be operating on the generalization mentioned above: All lexical causatives are transitive. In semantic terms, the learner knows that she or he is expressing a causative event. If in addition, the learner is working under this assumption that all causatives are syntactically transitive, their transitive use in the attested errors would follow.

While the latter approach would account for the learner transitivization errors and their correspondence to causativity without imputing to the learner any prior assumptions about verb alternations, what still remains striking is that among the 260 NV miscollocations, in all of the cases where the learner's choice of incorrect verb differs from the corresponding correct verb in transitivity, the intransitive and transitive verb counterparts also correspond to each other semantically, forming a causative-non-causative counterpart pair.   Were the learners simply misusing a verb, giving it an unwarranted causative meaning, what accounts for the fact that the verb they reach for is the non-causative counterpart of the causative they have invented?

The search for the optimum account of this data raises a more fundamental question of what our basic generalization is about, and in turn, about the nature of interlanguages. Recall that the generalization is that for each syntactic misuse of a verb with respect to transitivity (mostly the use of an intransitive verb as transitive), the incorrectly used verb is a non-causative that is semantically entailed in the meaning of the correct counterpart causative transitive verb that could replace the misused verb. Notice, however, that while the counterpart verbs in the verb pairs such as *appear* vs *show*, *remain* vs *keep*, *arise* vs *arouse*, etc. stand in just such a regular correspondence to each other, there is no particular reason, apart from this learner data, that we should notice these specific pairings in the English lexicon alone. What brings these verbs to our attention as pairs is not the target language per se, but learners' errors, for example, the misuse of the intransitive verb *appear* rather than a transitive such as *show*.

6. CONCLUSION

In analyzing a range of non-target VN co-occurrences extracted from a corpus of written English produced by learners in Taiwan, we have found two complementary sorts of lexical knowledge relevant: collocation and argument linking.

On the one hand, miscollocations are violations of conventionalized restrictions on co-occurrences of specific lexical items. Within these idiosyncractic preferences exhibited by specific lexemes, we have uncovered generalizations. Specifically, in VN collocations, it is overwhelmingly the noun that exerts its preferences over the verb, with noun standing as focal word and verb as collocate. Argument linking, on the other hand, shows the reverse direction, with the verb determining the selection and distribution of the noun, usually the head of an internal argument of that verb. Further, unlike collocation, argument linking does not impose idiosyncratic preferences for individual nouns, but selects classes of arguments.

In addition, we have suggested that collocation knowledge poses a surprising and

non-trivial puzzle for learnability. It appears that collocation knowledge is not attributable to simple exposure to co-occurrences of collocating lexemes in the linguistic input since collocation knowledge entails not only the realization that particular co-occurrences are acceptable, but also knowing that near synonyms are correspondingly unacceptable as replacements for the collocate. In the absence of negative evidence, it is not clear how this aspect of collocation knowledge could arise from the linguistic input either in L1 or L2 acquisition. This learnability puzzle deserves further research.

The learners' non-target cases of argument linking also touch upon acquisition puzzles, specifically concerning the origins of transitivity errors. First, learners' uses of strictly intransitive non-causative verbs as transitive causatives presumably would have no exemplars in the target language input that could account for their presence in the learners' production. Second, Chinese, the L1 of these learners, does not allow comparable causative transitives, thus ruling out a transfer account. Unlike research on verb alternations in L2 (Montrul 1997; Juffs 1996), we suggest attributing these over-transitivization cases to a simple default assumption on the part of the learners that semantically causative verbs are syntactically transitive. Our analysis suggests then the prospect that L2 lexical acquisition makes reference to a level of event structure, where the semantics of causative events can be given a unified transparent representation and can insure the predictable lexicalization of causative verbs as transitives. This syntax-lexical semantics interface makes it possible to capture the universal entailment that verbs which are lexicalized as semantically causative are expressed syntactically as transitives. Hence the learner's misconstrual of the semantics of the verb could account for the corresponding transitivization errors. On this view, a crucial locus of L2 lexical knowledge is event structure, and the hypothesis space of the learners is plausibly constrained by universal properties of event structure. In this paper we hope to have offered a narrow range of cases motivating such a mediation.

[1] The methodology is sketched in Wible, Kuo, Chien, Liu, and Tsao (2001) and detailed in Wible, Kuo, Tsao, Liu, and Lin (2001).

[2] For this sampling, we selected 100 nouns from LDEC by taking for each letter of the alphabet the first ten nouns listed which showed VN collocations (i.e., we skipped nouns that listed no VN collocations). We did this for the first ten letters of the alphabet to get our sampling of 100 nouns that participate in VN collocations.

[3] We borrow the term focal word from Manning and Schutze (1999:151-189) and the discussion there on collocation.

[4] Aside from our intuitions concerning the causativity component of the relevant verbs here, we find confirmation in Dorr's LCS Verb Database (LVD) http://www.umiacs.umd.edu/~bonnie/verbs-English.lcs, which is an extension and elaboration of Levin's English Verb Classes and Alternations (Levin 1993). All of the intransitive verbs from our Table 4 which the learners misuse with respect to transitivity have an entry in Dorr's LVD with an LCS (Lexical Conceptual Structure) representation that lacks the causative primitive ('cause'). Conversely, the correct transitive verbs that could replace these verbs all have entries with LCSs which do include the causative primitive. In the two reverse cases where the learners use a transitive verb intransitively ('shed' and 'cause'), the correspondence still holds. Those two transitives are listed with the causative primitive in the LCS representation, and the correct intransitives that could replace them have entries that lack the causative primitive in their LCSs

[5] Montrul (1997) inter alia pursues this line of research in second language lexical acquisition.

[6] The predicate decomposition analysis that we are suggesting could perhaps translate trivially into an approach using meaning postulates instead (Carnap 1947; Fodor 1987) though we do not pursue the possibility here.   See Chierchia and McConnell-Ginet (1990:360-66) for a comparison of the two approaches to lexical semantics.

[7] We point out in section 5.1 below that Chinese apparently has no morphologically simple causative verbs. Causatives are periphrastic in Chinese, created with special causative morphemes or with resultative complements added to the lexical verb. This, however, does not contradict a putative universal that lexical causatives, where they do exist in languages, are transitive. As we point out in 5.1, the absence of simple lexical causatives in Chinese rules out a straightforward transfer account of the 'over-transitivization' errors that we are considering in this section (See Tables 4 and 5).

[8] There is one verb in Table 4, however, which prima facia poses problems for this view. In the sentence *Lots of aftershocks caused constantly for several days*, we see the quintessential causative verb being used intransitively by this learner. This is an apparent counterexample to our suggestion that the learners assume that semantically causative verbs are syntactically transitive. We suggest the challenge is only apparent, however. First we need to be careful to distinguish the English lexeme 'cause' from the semantic primitive 'cause,' which should be taken as accidental homonyms here. The semantic primitive is expressed in a semantic metalanguage that resembles English only for expository ease. The point is that it is quite plausible that a learner in some stage of acquiring the English verb 'cause' may not grasp it as semantically 'causative', more plausible than appearances would lead us to believe. From the

context of this learner's error, for example, it is clear that it is being used to mean something more like '*Aftershocks occurred*' since there is no expression (even elided) of any result or change that the aftershocks are being claimed to effect, a defining property of causatives.

[9] Montrul (1997 and 2001) and Juffs (1996) make use of predicate decomposition and event structure representations in analyzing transitivity alternations in the English of L2 learners, but they look at the learners' acquisition of causative/inchoative alternations rather than simply at the transitive use of strictly intransitive verbs. Moreover, these studies use elicitation and judgment tasks and are therefore able to uncover not simply whether learners accept or produce causative transitives, but whether they accept alternations of particular verbs. Our data, extracted from a corpus of unrestricted texts produced by learners, contains instances where the learners produce the constructions unsolicited and in context (and this has its advantages). One consequence is that we focus simply on the transitive use of strictly intransitive verbs and not on learners' grasp of transitivity alternations. As we show below, however, this yields to a rather simple hypothesis.

[10] One important research question that we set aside here is whether the work done by argument structure representations such as PAS can be subsumed completely under an event structure level of representation such as LCS. While a number of linguists assume that a grammar includes both a PAS and an event structure representation (the latter feeding the former), Jackendoff (1990:46-49) proposes that LCS representations render PAS as well as any reference to thematic roles or hierarchies superfluous. Our data calls for a level of lexical decomposition such as that which an LCS or event structure would encode and needs no reference to predicate-argument structure or thematic roles or hierarchies. Thus it is consistent with the suggestion to eliminate an independent PAS and reference to thematic roles.

[11] We have subsequently found that Juffs (1996) makes this same observation about Chinese.

[12] Montrul (2001) uses elicited production tasks and learner grammaticality judgments to test specifically their perceptions of whether verbs of various classes (unaccusative, unergative) allow transitivity alternations.

REFERENCES

Bach, Emmon (1986) "The Algebra of Events," Linguistics and Philosophy 9:5-16.

Bazell, C., Catford, J.C., Halliday, M.A.K., and Robins, R.H. (eds.) (1966) *In Memory of J.R. Firth*. London: Longmans.

Benson, M. (1990) "Collocations and General Purpose Dictionaries," *International Journal of Lexicography* 3:1 pp. 23-35.

Bley-Vroman, R. (1983) "The Comparative Fallacy in Interlanguage Studies: The Case of Systematicity." *Language Learning* 33: 1-17.

Bowerman, Melissa (1990) "Mapping Thematic Roles onto Syntactic Functions: Are Children Helped by Innate Linking Rules?" *Linguistics* 28, pp 1253-1289.

Bresnan, J. (1982a) "Control and Complementation," *Linguistic Inquiry* 13: 343-434.

Bresnan, J. (1982b) *The Mental Representation of Grammatical Relations*. MIT Press: Cambridge, Massachusetts.

Croft, William (1998) "Event Structure in Argument Linking," in M. Butt and W. Geuder (eds) *The Projection of Arguments: Lexical and Compositional Factors.* Stanford, CA:CSLI.

Dowty, David (1979) *Word Meaning and Montague Grammar*. Reidel: Dordrecht.

Dowty, David (1982) "Grammatical Relations and Montague Grammar," Jacobsen and Pullum 1982.

Dowty, David (1991) "Thematic Proto-Roles and Argument Selection," *Language* 67: 547-619.

Filip, Hana (1994) "Aspect and the Semantics of Noun Phrases," in Co Vet and Carl Vetters (eds.) *Tense and Aspect in Discourse*. New York: Mouton.

Firth, J.R. (1957) A Synopsis of Linguistic Theory 1930-1955. In Studies in Linguistic Analsysis, pp.1-32. Oxford: Philological Society. Reprinted in F.R. Palmer (ed.), Selected Papers of J.R. Firth 1952-1959, London: Longman, 1968.

Fodor, Jerry (1970) "Three reasons for not deriving *kill* from *cause to die*," *Linguistic Inquiry* 1:429-38.

Goldberg, Adele (1995) *A Construction Grammar Approach to Argument Structure.* Chicago: University of Chicago Press.

Grimshaw, J. (1990) *Argument Structure*. Cambridge, Mass: MIT Press.

Halliday, M.A.K. (1961) "Categories of the Theory of Grammar," *Word* 17:241-92.

Halliday, M.A.K. (1966) "Lexis as a linguistic level," In C.E. Bazell, J.C. Catford, M.A.K. Halliday, and R.H. Robins (eds.) In *Memory of J.R. Firth*, pp 148-162, London: Longmans.

Hinrich, Erhard (1985) *A Compositional Semantics for Aktionsarten and NP Reference in English*. Doctoral dissertation, Ohio State University.

Jackendoff, R. 1990 *Semantic Structures*. Cambridge, MA: MIT Press.

Juffs, A. (1996) *Learnability and the Lexicon: Theories and Second Language Acquisition Research*. John Benjamins: Amsterdam.

Kilgarriff, A. and D. Tugwell (2001) "WORDSKETCH: Extraction, Combination, and Display of Significant Collocations for Lexicography." *Workshop Proceedings: Collocation: Computational Extraction, Analysis and Exploitation.* 39th Meeting of the Association for Computational Linguistics, Toulouse, France.

Krenn, B. (2000) *The Usual Suspects: Data-Oriented Models for the Identification and Representational Models for the Identification and Representation of Lexical Collocations*. DFKI & Universitat des Saarlandes, Saarbrucken.

Krifka, Manfred (1992) "Thematic Relations as Links between Nominal Reference and Temporal Constitution," in I. Sag and A. Szabolsci (eds.) *Lexical Matters*, CSLI, Stanford.

Kuo, Chin-Hwa, David Wible, and Nai-Lung Tsao (2001) "On the Design of a Web-based English Writing Environment and Learner Corpus," *IEICE Transactions on Information and Systems.* Vol E84-D, pp 1-10.

Lakshmanan, U. and L. Selinker (2001) "Analysing Interlanguage: How do we Know What Learners Know?" *Second Language Research* 17:4, pp. 393-420.

Lasnik, H. (2000) *Syntactic structures revisited.* Cambridge, MA: MIT Press.

Levin, B. (1993) *English Verb Classes and Alternations: A Preliminary Investigation.* Chicago, IL: University of Chicago Press.

Lewis, M. (2000) *Teaching Collocation: Further Developments in the Lexical Approach.* Language Teaching Publications, London, England.

Lewis, M. (2000) "Language in the lexical approach." In *Teaching Collocation*, Lewis, M., ed., Language Teaching Publications, London, England. pp. 132-133.

Levin, B (1995a) *Unaccusativity: At the Syntax-Lexical Semantics Interface.* Cambridge, Mass: MIT Press.

Liu, A. (2002) *A Corpus-based Lexical Semantic Investigation of Verb-Noun Miscollocations in Taiwan Learners' English.* MA thesis, Tamkang University, Taipei.

Manning, C.D. and H. Schutze (1999) *Foundations of Statistical Natural Language Processing.* Cambridge, MA: MIT Press.

Marantz, A. (1984) *On the Nature of Grammatical Relations.* MIT Press, Cambridge, Massachusetts.

Montrul, Silvina (2001) "Causatives and Transitivity in L2 English," *Language Learning* 5:1 pp. 51-106.

Montrul, Silvina (2000) "Transitivity Alternations in L2 Acquisition," Studies in Second Language Acquisition 22:229-273.

Montrul, Silvina (1997) *Transitivity Alternations in Second Language Acquisition: A Cross-Linguistic Study of English, Spanish, and Turkish.* Ph.D. dissertation, McGill University.

Nakhimovsky, A.D. and Leed, R.L. (1979) "Lexical functions and language learning," *Slavic and East European Journal*, 23:1.

Oshita, H. (1997) *The Unaccusative Trap: L2 Acquisition of English Intransitive Verbs.* Doctoral dissertation, University of Southern California.

Perlmutter, D (1978) "Impersonal Passives and the Unaccusative Hypothesis," In *Proceedings of the Fourth Annual Meeting of the Berkeley Linguistics Society.* Pp 157-89. Berkeley Linguistics Society, University of California, Berkeley

Pesetsky, D. (1982) *Paths and Categories.* Doctoral dissertation, MIT, Cambridge, Mass.

Pinker, Steven (1989) *Learnability and Cognition: The Acquisition of Argument Structure.* Cambridge, MA: MIT Press.

Pinker, Steven (1999) *Words and Rules.* Cambridge, MA: Basic Books.

Pustejovsky, J (1995) *The Generative Lexicon.* Cambridge, MA: MIT Press.

Rappaport, M., M. Laughren, and B. Levin (1993) "Levels of Lexical Representation," in J. Pustejovsky (ed.) *Semantics and the Lexicon* Kluwer: The Netherlands.

Rappaport Hovav, M. and Levin, B. (2001) "An Event Structure Account of English Resultatives," *Language* 77:766-797.

Rappaport Hovav, M. and B. Levin (1998) "Building Verb Meanings," in M. Butt and W. Geuder (eds.) *The Projection of Arguments: Lexical and Compositional Factors.* CSLI: Stanford.

Rutherford, W.E. (1987) *Second Language Grammar: Learning and Teaching.* Harlow: Longman.

Shibatani, Masayoshi (1976) "The grammar of causative constructions: A conspectus," In M. Shibatani (ed.) *The Grammar of Causative Constructions* (*Syntax and Semantics*, vol. 6) New York: Academic Press.

Shimohata, S. (1997) "Retrieving Collocations by Co-occurrences and Word Order Constraints," in Proceedings of ACL-EACL '97, pp. 476-481.

Sinclair, J.M. (1966) "Beginning the study of lexis" in Bazell, C. et al. (eds.) pp. 410-430.

Sinclair, J.M. (1991) *Corpus, Concordance, and Collocation.* Oxford: Oxford University Press.

Smadja, Frank (1993) "Retrieving collocations from text: **Xtract,**" Computational Linguistics 19:1, pp 143-177.

Talmy, L. (1980) "Lexicalization Patterns: Semantic Structure in Lexical Form," In T. Shopen et al (ed.) *Language Typology and Syntactic Description*, vol. 3 New York: Cambridge University Press.

Tenny, C. and J. Pustejovsky (2000) "A history of events in linguistic theory," in C. Tenny and J. Pustejovsky (eds) *Events as grammatical objects.* Stanford, CA: CSLI Publications.

Wible, David, Chin-Hwa Kuo, Nai-Lung Tsao, Anne Liu, Hsiu-Ling Lin (2003) "Bootstrapping in a Language Learning Environment," *Journal of Computer-Assisted Learning*, vol 19 #1, pp. 90-102.

Wible, David, Chin-Hwa Kuo, Feng-yi Chien, Anne Liu, and Nai-Lung Tsao (to appear) "A Web-based EFL Writing Environment: Exploiting Information for Learners, Teachers, and Researchers," *Computers and Education,* 37:297-315.

Wible, David, Chin-Hwa Kuo, Nai-Lung Tsao, and Anne Liu (2002) "On Finding a Computable Path from Miscollocations to Collocations." Unpublished manuscript, Tamkang University, Taipei

Williams, E. (1981) "Argument Structure and Morphology," *Linguistic Review* 1:81-114.

Woolard G. (2000) "Collocation: encouraging learner independence*,"* In *Teaching Collocation*, Lewis, M., ed., Language Teaching Publications, London, England.

Yarowsky, David (1993) "One Sense per Collocation," In *Proceedings of ARPA Human Language Technology Workshop*, pp. 266-271. Princeton, NJ.

Yip, V. (1994) "Grammatical Consiousness-raising and Learnability," in Terence Odlin (ed.) Perspectives on Pedagogical Grammar. Cambridge: Cambridge University Press.

Yip, V. (1995) *Interlanguage and Learnability: From Chinese to English.* Amsterdam: John Benjamins.