

GRAMMAR AND COGNITION IN SINITIC NOUN CLASSIFIER SYSTEMS

James Myers

Jane Tsay

National Chung Cheng University

Abstract. This paper is a brief summary of some of our recent research on the processing of nominal classifiers in two Sinitic languages spoken in Taiwan, Mandarin and Taiwanese (Southern Min Chinese). Our research has used data from classifier systems to address two interface issues, namely the relation between grammar and extra-grammatical cognition and between grammar and the lexicon. We have been researching these questions with several different theoretical and methodological approaches: descriptive linguistics, experiments with adults, longitudinal records of child language acquisition, and connectionist modeling.

Keywords: classifiers, grammar vs. cognition, grammar vs. lexicon

1. Introduction

This paper is a brief summary of some of our recent research on the processing of nominal classifiers in two Sinitic languages spoken in Taiwan, Mandarin and Taiwanese (Southern Min Chinese).

Nominal classification systems are rather common in languages across the world (Allan 1977, Aikhenvald 2000). Languages beyond the Sinitic family that have noun classifiers include Japanese, Thai, and many other languages of East Asia; Swahili, Tiv and many other languages of Africa; Navajo, Ojibwa, Inuit, and many other languages of the Americas; and many languages of the Pacific and Australia. In fact, the only major language family where grammatical noun classifiers are not found is the Indo-European family, though some have claimed that gender behaves like an extremely impoverished classification system.

Classifier systems have excited much interest among cognitive linguists, not only because so many human brains across the planet process classifiers as part of their everyday business, but also because classifier systems provide a very fruitful source of data on two questions fundamental to the understanding of the cognitive nature of human language.

First, classifiers represent a link between grammar and extra-grammatical cognition. By definition, classifier languages have grammars that require the use of special morphemes in certain syntactic contexts. However, as the term "classifier" implies, the choice of classifier for a particular noun is not determined solely by memorization or by a simple grammatical rule, but rather is in principle subject to all the complex cognitive factors that go into human categorization of entities in the real world. This fact means that classifier

systems can be used to address an important question: Is the structure and processing of language essentially autonomous of extra-grammatical cognition (e.g. Pinker 1994, Newmeyer 1998), or are there strong causal relations between language and general cognition (e.g. Lakoff 1986, Langacker 1991)?

The second question on which classifiers can shed some light is equally fundamental: Does the knowledge of language essentially consist of knowledge of general, symbol-manipulating rules (e.g. Pinker 1991, 1999) or can so-called grammatical behavior be explained entirely by a powerful associative lexicon (e.g. Bates and Goodman 1997)? Classifier systems can provide data for this debate because classifiers are often used in conventionalized ways that seem to require storing information in the lexicon, and yet classifier systems typically have a "default" classifier, apparently used solely to fulfill grammatical requirements.

We have been researching these questions in Mandarin and Taiwanese with several different theoretical and methodological approaches. In this paper, we first describe some descriptive linguistic work and experiments that address the question of grammar versus extra-grammatical cognition. Next we turn to experimental studies that address the question of grammar versus the lexicon. Then we describe some relevant results from an on-going study of the acquisition of Taiwanese classifiers. Finally, we look at some connectionist modeling we have been doing to address both of these questions.

2. Grammar and cognition

We begin with a description of the Mandarin and Taiwanese classifier systems, both to introduce these systems and to point out how they illustrate the interaction between grammar and extra-grammatical cognition.

In the Sinitic languages, a classifier is a monosyllabic morpheme that is required in noun phrases containing determiners or numbers (e.g. Li and Thompson 1981; Tai 1992, 1994). However, Sinitic classifiers do not all behave the same way. At one extreme there are what have been called measure words (see e.g. Tai 1992, 1994 for recent discussions) or massifiers (Cheng and Sybesma 1998), which primarily serve the function of quantification rather than classification. An example is Mandarin *bang* "pound", which appears in structures quite parallel to those found in European languages, such as *yi-bang rou*, literally "a pound meat". One can also say *yi-bang DE rou*, literally "a pound OF meat".

At the other extreme are what have been called individual measures (Chao 1968) or individual classifiers (e.g. Ahrens and Huang 1996), which appear to be selected by individual entities on the basis of their inherent semantics. These have no parallel in European languages, and unlike the case with standard measures, noun phrases with individual measures cannot include the morpheme *de*. An example in Mandarin is *tiao*, which tends to be used for oblong objects of certain kinds, as in *yi-tiao lu* "a road" or *yi-tiao yu* "a fish" (see e.g. Tai and Wang 1990).

Nevertheless, there is no sharp line between classifiers that quantify and those that classify. In between standard measures and individual measures are container measures (such as *bei* "cup"), group measures (such as *qun* "group"), and partitive measures (such as *kuai* "piece"). These vary in their syntactic and semantic properties, and in fact some classifiers belong to more than one type, such as *pian*, which can either be a partitive measure

meaning "slice", or an individual measure for flat objects like leaves. Such observations alone suggest that it is not possible in general to draw a sharp line between grammar and extra-grammatical cognition.

Further evidence for this view comes from a closer examination of the individual classifiers. Besides *tiao*, other Mandarin classifiers include *zhang*, *zhi*, *wei*, and *liang*. *Zhang* tends to be used for objects with a two-dimensional extended surface, such as *yi-zhang zhi* "a piece of paper" or *yi-zhang zhuozi* "a table" (e.g. Tai and Chao 1994). *Zhi* is most often used for animals, such as *yi-zhi gou* "a dog" or *yi-zhi houzi* "a monkey". *Wei* is used for people to whom one should be polite, for example *yi-wei laoshi* "a teacher"; it would be strange to say *yi-wei xiaotou* "a thief". Finally, *liang* is used for vehicles, such as *yi-liang qiche* "an automobile." The properties used to define the categories thus include shape, animacy, humanness, and function, which are precisely those that play central roles in human cognition in general (Tai 1994).

The Taiwanese system is different from Mandarin in subtle but interesting ways. For example, in Mandarin, "fish" takes *tiao*, but in Taiwanese, the cognate *tiau* is only used for inanimate objects. Taiwanese speakers instead prefer to classify fish with *bue*, which literally means "tail". Here the Taiwanese speakers use the part of an entity to represent the whole, which is also a common strategy in human categorization (Tai 1994).

The different ways by which nouns select their classifiers further illustrates how difficult it is to separate grammar from "non-grammar". In some cases, classifiers are clearly selected for purely linguistic reasons, while in other cases, the deciding factor appears to be the actual observable properties of the physical entity. The Mandarin classifier *tiao* presents an interesting example of this. In certain domains, speakers seem to choose *tiao* because an object is actually oblong. For example, in Mandarin there are many vegetables and fruits whose compound names end in the root *gua*, such as *sigua*, literally "silk melon", which is a kind of vegetable called "loofah" or "towel gourd" in English, and *xigua*, literally "Western melon", which is the Chinese word for "watermelon". For the oblong *sigua*, the preferred classifier is *tiao*, while for the usually rounder *xigua*, *tiao* is not preferred (Tai and Wang 1990). Thus the fact that these compounds share a head morpheme plays no role in classifier selection. However, sometimes *tiao* seems to be preferred for a noun simply because that noun is linguistically similar to the name of an oblong object. For example, in Mandarin, *kuzi* "pants" usually takes the classifier *tiao*, presumably because pants are oblong. Nevertheless, *duanku* "shorts" are by definition short, and yet they also tend to take *tiao*, presumably because this word shares the root morpheme *ku* with *kuzi* (Wiebusch 1995).

Similarly, *tiao* is the preferred classifier for fish in Mandarin, even if the fish are not oblong. A flounder, for example, is classified with *tiao*, rather than the "flat" classifier *zhang*. This may also be due to linguistic convention, but a more interesting possibility is that sensitivity to shape is overridden by the fact that fish are animate. There seems to be a tendency that shape plays a more important role in the categorization of inanimate objects than of animate objects, as shown for example by studies of word learning (Landau 1996).

A final piece of evidence that classifiers are not selected by simple semantic rules of grammar comes from the observation that classifiers show prototype effects. That is, in a class of nouns where some classifier tends to be preferred, not all of the nouns are created equal: the classifier will be more preferable for some nouns compared to others. This has been observed for Mandarin by several researchers (e.g. Ahrens 1994, Tien 1996,

Chang-Smith 2000 for Mandarin, Tai 1997 for Taiwanese).

In one of the studies reported in Tai (1997), monolingual speakers of Taiwanese and bilingual Taiwanese/Mandarin speakers were shown pairs of pictures and then were asked to compare the pictures in words. Since the pictures in each pair showed different numbers of the same kinds of object, the grammar of Taiwanese forced the speakers to use classifiers. The question of interest was how varying properties of the objects would elicit different classifiers. Table 1 shows the results for the classifier *bue* "tail".

Table 1: Proportions of *bue* responses by adult speakers of Taiwanese

	snake	fish	dragon	shrimp	lizard	dolphin
monolingual (Taiwanese only)	100%	93%	75%	75%	58%	42%
bilingual (Taiwanese/Mandarin)	83%	68%	35%	65%	18%	40%

As expected, *bue* was overwhelmingly the preferred classifier for snakes and fish; these entities thus seem to be prototypes for the *bue* category. The likelihood that other animals receive this classifier then seems to be a function of how similar they are to the prototypes. Thus among both monolingual and bilingual speakers, over half used *bue* for shrimp, which are cold-blooded animals like snakes and fish, but less than half used *bue* for dolphins, which are mammals. The choice of classifier for dragons and lizards, which are snake-like but have legs, varied dramatically between monolingual and bilingual speakers: the former still preferred *bue*, but speakers who also spoke Mandarin preferred *ciah*, the Taiwanese cognate of *zhi*, which Mandarin can use for all four-legged animals.

3. Grammar and the lexicon

Even considered as linguistic entities, however, classifiers have a dual existence, midway between the lexicon and grammar proper. One sign of the lexical nature of classifiers is shown by the prototype effects noted earlier. That is, when speakers are deciding whether to use a classifier for an object, they seem to do this by comparing the object with the classifier category's prototypical member. Thus there is no grammatical rule in Taiwanese that says "*Bue* must be used for entities that have tails, are alive, live in the water, are cold-blooded, and have no legs." Instead dolphins, for instance, are compared with an internal mental representation of fish, and found to be similar, but not as similar as, say, shrimp.

Another example of the lexical nature of classifiers is shown by the contrast between the oblong classifier *tiao* in Mandarin and two other classifiers also used for oblong objects, *gen* and *zhi* (homophonous with the animal classifier). There do seem to be some semantic generalizations in how these three classifiers are used. Thus *tiao* tends to be used for oblong flexible objects like fish and pants, *gen* for oblong rigid objects like bananas and fingers, and *zhi* for oblong, rigid, cylindrical objects like pens (Tai and Wang 1990).

However, each of these generalizations has so many exceptions that it doesn't seem likely that speakers are really following semantic rules when they choose which of these three classifiers to use. Thus cucumbers are not flexible but take *tiao*, ropes are flexible but can take *gen* as well as *tiao*, and ice cream bars aren't cylindrical but they take *zhi*. On top of this, classifiers may be used in purely conventionalized ways that simply must be memorized.

For example, the word for "law" takes *tiao*, because historically laws were written on long strips, and even though towels are both oblong and flat, towels take *tiao*, never the "flat" classifier *zhang*. Therefore, it appears likely that speakers memorize common exemplars for each classifier and then generalize from them by a kind of exemplar-driven analogy.

However, both Mandarin and Taiwanese also have a so-called general classifier, or default classifier, pronounced *ge* in Mandarin and *e* in Taiwanese. In both languages, the default classifier is by far the most commonly used, though *ge* is used more often in Mandarin (up to 95% of all classifier tokens in a study of spoken Mandarin discussed in Erbaugh 1986) than *e* is used in Taiwanese (approximately 50% of all classifier tokens according to our own estimates from radio broadcasts). Since classifiers are required by the syntax of these languages, speakers will often use these apparently meaningless defaults merely to fulfill the syntactic obligations. This reveals an important non-functional aspect to classifier systems, since using default classifiers so often is as useless for classificatory purposes as stuffing most office papers into a huge "miscellaneous" file.

The Mandarin and Taiwanese default classifiers are defaults in all three senses discussed in the cross-linguistic survey of Zubin and Shimojo (1993) (though their point there was to argue that the three senses do not necessarily always cooccur, e.g. in Japanese). First, they may be used for nouns that have no other classifiers (see also Loke 1994). These include names for people that you don't have to be polite to, such as thieves; objects that don't have any particular shape, such as roundish watermelons; abstractions, such as "nation"; and nouns derived from verbs, such as "dream". This gap-filling job of a default classifier thus results in a class of nouns that is semantically incoherent, or at best, one could say that the semantics of *ge* are extremely ambiguous.

Second, the default classifier can replace other classifiers. For example, in Mandarin, a table, which normally takes the "flat" classifier *zhang*, is sometimes also used with *ge* in informal speech.

The tendency to neutralize to *ge* is affected by a number of factors, all of which reemphasize the importance of nongrammatical cognitive factors in classifier selection, even when speakers may merely be attempting to fulfill syntactic requirements of noun phrase structure. One such factor concerns how similar a noun is to the prototype for its classifier category. Thus Ahrens (1994) found that paper, which is of course extremely flat, is far more likely to take *zhang* than *ge*, while a bed, which has a flat surface but is also three-dimensional, often takes *ge* instead of the prescriptively correct *zhang*. A second factor affecting neutralization to *ge* is the type of semantic information marked by the classifier. Loke (1994) observes that *ge* is more likely to replace a classifier marking function than one marking shape or animacy. Yet another factor is informativeness; Erbaugh (2000) found that Mandarin speakers tend to use a specific (i.e. non-*ge*) classifier for a noun the first time it's mentioned in a discourse, while later references tend to use the general classifier or a bare noun (no classifier). Nevertheless, in principle *ge* can replace all individual classifiers, and there are even some speakers of Mandarin who rarely use other classifiers at all (Erbaugh 1986). The behavior of the Taiwanese default classifier *e* has not been as well studied, but it also appears that neutralization to *e* is not rare in Taiwanese either.

The final typological property of default classifiers in Mandarin and Taiwanese is that they can be used in contexts where the meaning of an entity is extremely vague. Thus in both languages, speakers can use the morpheme for "that" plus *ge* or *e* to refer to any kind of

entity when the name for the entity itself is not used, as in Mandarin *Na-ge shi shenme?* "What is that?"

Together these three properties are consistent with the hypothesis that default classifiers are selected without the use of exemplar-driven analogy. Indeed, the only thing all of these phenomena have in common is that they all involve cases where analogy fails, either because there are no relevant exemplars, or because the memory traces or cognitive salience for potential exemplars is too weak. This view predicts that speakers with memory access problems will tend to overuse default classifiers, and this is in fact true: at least for Mandarin, it has been shown that dysphasic speakers (Tzeng, Chen and Hung 1991), nonnative learners of Chinese (Polio 1994), and children whose lexicons are still growing (Erbaugh 1986) all overuse the default classifier *ge*. Further evidence from children acquiring Taiwanese will be discussed in the next section.

We therefore have been struck by the parallels between the default classifier in the Sinitic languages and regular inflection in languages like English (see Myers 2000 for a detailed exposition of these parallels). Steven Pinker and others have argued that regular inflection is processed by a grammatical module quite distinct from the associationist lexicon that processes irregular inflection (Pinker 1991, 1999), while numerous connectionists have tried to show that an associationist lexicon is sufficient to handle all kinds of inflection (e.g. Rumelhart and McClelland 1986, Plunkett and Marchman 1993). In our attempts to determine which perspective in this debate is more appropriate for classifier systems, we have conducted a set of experiments on adult speakers of Mandarin.

Our experiments have been of two types. In the first type, we have tested the hypothesis that as a grammatical default, *ge* should have no semantic properties of its own (Myers, Gong and Shen 1999). One experiment of this sort was adapted from research on regular inflection by Steven Pinker and colleagues (e.g. Prasada and Pinker 1993, Marcus, Brinkmann, Clahsen, Wiese, and Pinker 1995). These earlier experiments tested the claim that speakers process regular inflection independently of exemplars in memory. Prasada and Pinker (1993), for example, asked whether experimental participants could produce regular past tense forms as easily for nonsense words that are phonologically different from real regular verbs as for nonsense words that are similar to real regular verbs. For example, is it as easy to change strange-sounding *ploamph* into *ploamphed* as it is to change *plip*, which is similar to real words like *slip*, into *plipped*? As expected by the grammatical view of regular inflection, no difference was found between these two kinds of nonsense forms.

In our case, we gave experimental participants nonsense forms with experimentally manipulated "meanings", some quite similar to words that typically take *ge* (such as nouns for humans) and others not similar to any specific lexical items (such nouns that were given no semantics at all). If *ge* is selected by exemplar-driven analogy, we might expect that speakers would be more likely to choose *ge* for the "human" nouns than for the meaningless ones.

Meanings were given by defining a two-syllable nonword in terms of real words. Two conditions were particularly important. In the HUMAN condition, the nonwords were defined by referring to words for humans that tend to cooccur with *ge* (and not *wei*), and in the VAGUE condition, the nonwords were given no meaning at all. The task was simply to fill in a gap in a sentence with a classifier. For example, in the HUMAN condition, a participant might hear the following in Chinese: "A *da'nan*³ is a kind of bad guy; here there

are many *da'nan*³, over there are five [GAP] *da'nan*³." In the VAGUE condition, they might hear: "Here there are many *da'nan*³, over there are five [GAP] *da'nan*³."

The response patterns of our Mandarin-speaking speakers were quite consistent across all conditions. The results for the HUMAN and VAGUE conditions were especially striking: the proportions of *ge* responses were statistically indistinguishable. Thus, just as with regular inflection in English, it appears that speakers of Mandarin do not find it harder or more unnatural to choose the default classifier for words without lexical exemplars.

The second type of experiment is inspired by research on how lexical frequency affects the processing of inflection. If regular inflection is processed outside of the lexicon in some sense, then we might expect it to be insensitive to lexical frequency effects. Unfortunately, this is not strictly true. For example, Sereno and Jongman (1997) reported that regularly inflected plural nouns in English showed the standard frequency effects in lexical decision tasks: higher frequency plurals were responded to faster, regardless of the frequency of the associated singular. In a recent review of the literature, however, Pinker (1999) suggests that it depends on the task whether or not frequency affects the processing of regular inflection. Off-line tasks, such as making acceptability ratings of past tense forms (Ullman 1999), only show frequency effects for irregulars, not for regulars. This task-dependence is presumably due to the fact that experience of any kind leaves memory traces; under time pressure (as in a reaction-time experiment), it's always more efficient to simply retrieve information from memory rather than rederiving it. Without time pressure, however, other processes can come into play, and this is when regular inflection seems to be processed without much contact with the lexicon.

If *ge* is processed outside of the lexicon, as we hypothesize, we also expect it to be immune from a kind of frequency effect, at least for certain tasks. Specifically, it should not matter to speakers how often a noun has collocated with *ge* in their experience with the language. By contrast, ordinary classifiers like "oblong" *tiao* or "polite-human" *wei* should be affected by the collocation frequency between classifier and noun.

In our experiments testing these predictions, instead of pure collocation frequency, we have been using the measure of "mutual information", or MI, which is defined as the base-two logarithm of the collocation frequency of two words divided by the frequencies of each of the words (Church and Hanks 1990; Chen, Huang, Chang and Hsu 1996). MI, commonly used in corpus linguistics, can be thought of as a measure of how predictable one word is from another. Higher MI indicates words tend to collocate together more often than would be expected from the chance encounters caused merely by the high frequencies of one or both of the words. Thus if the results found with English inflection carry over to classifiers, we expect that the effect of MI should depend on the time pressure given to the experimental participants: without time pressure, there should be no MI effect for nouns that prefer the default *ge*, but with time pressure there perhaps may be such an effect.

In one of our experiments, Mandarin speakers were shown nouns that can appear with more than one classifier (e.g. *yu* "fish", which appears with both "oblong" *tiao* and "animal" *zhi*). Each classifier-noun collocation was labeled as more or less common using MI. Participants were shown each of the nouns and asked which of the two collocating classifiers was more appropriate; they made their choices on paper with no time pressure. Results showed an overall significant positive correlation between MI and classifier choice: the higher a classifier-noun MI, the more likely participants were to choose that classifier for that

noun. However, it turned out that this was solely due to nouns that can take both *ge* and "polite-human" *wei*. The correlation was significant both for nouns that take *ge* more often and for nouns that take *wei* more often.

It was quite a surprise that MI seemed to have no effect on judgments for most classifier-noun collocations, even among non-default classifiers. The proper interpretation of the overall lack of relevance of MI may in fact require reference to the extra-linguistic cognitive factors discussed earlier. That is, it may be that the only time speakers do rely on factors such as MI is when distinguishing nouns for "ordinary" people (who can take *ge*) from people one should be respectful to (taking *wei*); this difficult and somewhat arbitrary judgment apparently requires the assistance of non-cognitive factors such as MI. In any case, the results do not show that *ge* selection is directly affected by MI, since the task specifically contrasted it with *wei*, and it may be that it is *wei* that is affected by MI (e.g. if the MI between a given noun and *wei* is low, the speaker is forced to choose *ge* by default).

In another experiment, this time using time pressure, our goal was to see if differences in classifier-noun MI affect the speed with which classifiers are selected. We expected that it would have an effect for most classifiers, but not for *ge*. Materials were nouns that varied along two dimensions: classifier preference (i.e. *ge* or non-*ge*) and classifier-noun MI (e.g. a noun may prefer *ge* when it appears with a classifier, but usually it appears with no classifier at all). Since by definition MI correlates inversely with lexical frequency (the formula for MI involves dividing by the frequency of each collocating word), we could not vary MI and also control lexical frequency. Lexical frequency is of course known to have a strong effect on reaction time (RT), so this was a problem. We solved it by having each experimental participant perform two tasks: a "naming" task where participants simply read aloud an NP (written in Chinese characters) containing a classifier and noun (e.g. *san-tiao xiangjiao* "three-TIAO bananas") and a "production" task where participants filled in the classifier for a noun (e.g. *san-__ xiangjiao*). Both of these tasks are affected by lexical frequency, but presumably only the second is also affected by MI in classifier selection. We then subtracted each participant's RT for each item in the naming task from the production task to get a measure of the time required for classifier selection alone (the order of tasks was counterbalanced across participants).

The results were as follows. Since Mandarin allows some freedom in classifier selection, we first analyzed classifier choice. Unsurprisingly, participants often chose classifiers other than those prescriptively preferred (or those used in written Chinese). We discovered a significant difference in the behavior of *ge* vs. non-*ge* nouns: nouns that we expected to prefer *ge* were used with an alternative classifier at the same rate regardless of the MI between the noun and the expected *ge*, but nouns that we expected to prefer non-*ge* classifiers were used with an alternative much more often if MI was low. This seems to be strong evidence for the notion that selecting *ge* occurs independently of the lexicon, while the selection of other classifiers involves lexical factors such as MI.

Reaction times were also analyzed (i.e. production RT minus naming RT, presumably reflecting primarily RT for classifier selection). There were significant effects both for classifier preference (selecting the classifier for *ge* nouns was slower) and for MI (high MI nouns were slower), and there was a significant interaction. We also found that for the *ge* nouns, there was a positive correlation between RT and MI, but for the non-*ge* nouns there was no correlation.

The RT results are unfortunately very strange according to any model. First, the fact that high-MI nouns are slower makes no sense, since higher cooccurrence frequency should help speakers produce the correct classifier, not slow them down. At first this may seem to be a confound due to lexical frequency, which is inversely correlated with MI, but regression analysis found no correlation whatsoever between RT and lexical frequency for either the *ge* nouns or the non-*ge* nouns, so apparently our two-task method did indeed remove lexical frequency effects. Second, the fact that the interaction between classifier preference and MI is significant shows that *ge* and non-*ge* nouns behaved differently with respect to MI, as expected. However, the difference is the opposite of what was predicted: it was the *ge* nouns that were sensitive to MI, not the non-*ge* nouns. If the results are not an artifact, one possible explanation is similar to the analysis of the previously described experiment. That is, participants' choice between two non-*ge* classifiers is affected primarily by cognitive factors, but the choice between *ge* and non-*ge* classifiers is affected by MI. This in turn would presumably be due to the way non-*ge* classifiers are processed, so the MI effect is not directly due to *ge*.

In the most recent experiment conducted along these lines, we used only "human" nouns that either strongly prefer polite *wei* or strongly prefer neutral *ge*. We hoped that this would simplify the analysis somewhat, given that the off-line study only found MI effects for nouns of this type. Again we analyzed the two factors of MI and classifier choice. This time no effects were found for reaction time, but we did find that lower MI tended to increase error rate, that is using *wei* instead of *ge* or vice versa. However, there was no interaction between the factors of MI and classifier choice, which means that both *wei* and *ge* were affected equally by MI. This in turn may be because of the nature of the task (time-pressured), but it may also indicate that the processing even of the so-called default classifier *ge* is far more complex than it appears from a simple description of its role in the classifier system.

Summarizing this portion of our research, then, we have been examining whether default classifiers are processed independently of the lexicon. Overall results tend to point towards a positive answer, but we are willing to change our mind.

4. Language development

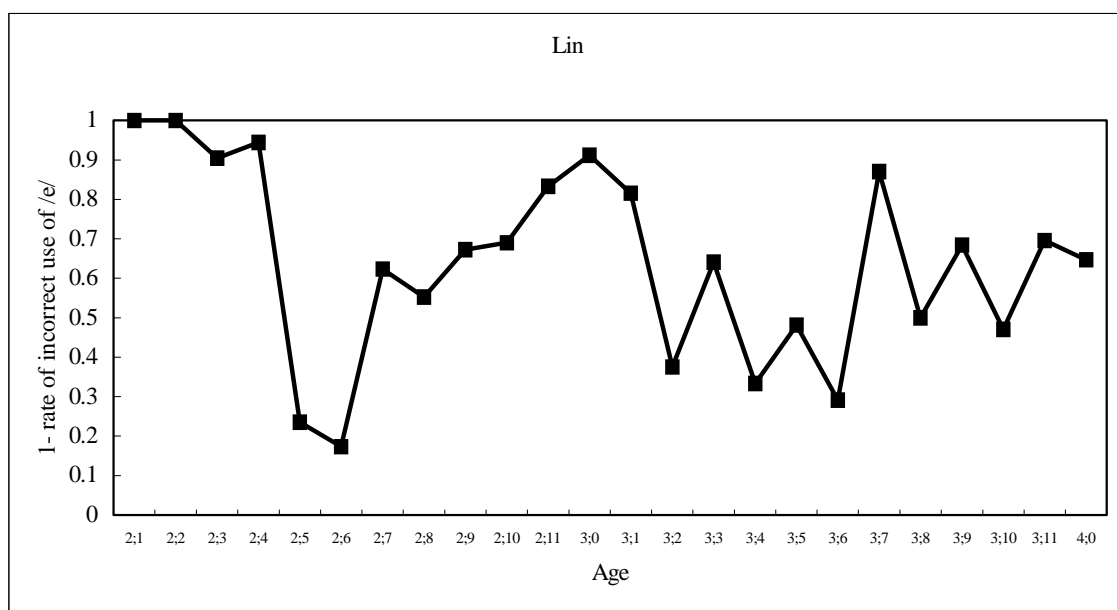
We have also been studying the interaction of grammar with cognition and with the lexicon using a very different line of evidence, namely from language development. Previous work on Mandarin and Taiwanese has found strong evidence for both cognitive and lexical effects in children's classifier productions (e.g. Erbaugh 1986, Hu 1993, Ng 1989). In our work, we have focused on two- to four-year-old children acquiring Taiwanese, since most work on classifier acquisition has examined older children. Our data come from a much larger project that involves building a corpus of spontaneous speech production from a set of children over the course of three years (Tsay 1997-2000; Tsay 2001-2003). While our analysis is still in progress, the data on Taiwanese classifier acquisition that we have looked at so far point to some interesting conclusions (see Myers and Tsay 2000 for further data and discussion).

First, the role of cognitive properties appears to be far more limited in the earliest use of classifiers than it is later on in life. This is consistent with what has been found for young children acquiring Thai classifiers and other noun categorization systems (Carpenter 1991). Thus it appears that in spite of the importance that cognitive factors have for processing classifiers by older children and adults, these factors do not drive initial classifier acquisition.

For example, one child, Lin, tended to neutralize even the animal classifier *ciah* to the default classifier *e*. This is in spite of the facts that animacy is a highly salient semantic property, and that studies on older children acquiring Taiwanese or Mandarin have found that children tend to overuse *ciah* or the Mandarin cognate *zhi*, even for inanimate objects (Hu 1993, Ng 1989). It may be, then, that at least for some young children, classifiers are learned as purely grammatical entities; cognitive salience does not affect their classifier choice as it does for older children and adults.

Our second tentative finding concerns the nature of the interaction between such lexical learning of classifiers and general grammatical rules. We have found that, like the acquisition of inflection (e.g. Marcus, Pinker, Ullman, Hollander, Rosen and Xu 1992), the development of the Taiwanese classifier system tends to follow a U-shaped learning curve. Thus at an early stage classifier use matches adults. Then around the age of two years six months, accuracy in classifier production drops dramatically, with almost all of the errors involving overuse of the default classifier *e*. Next, at least for some of the children studied, there is a gradual increase in accuracy over the remaining months for which we have data. For example, the following figure shows the accuracy curve for Lin (including all classifier productions, even in headless NPs).

Figure 1.



As with inflection, it is important to determine if this developmental pattern results from innate maturational processes or if it is driven by environmental factors. For example, is the U-shape a mere illusion resulting from the learning of new words for which the children don't know the classifier? The answer to this seems to be no. Sudden drops in accuracy followed by gradual improvement is also seen if we look at individual words. Another question that may be raised concerns whether the onset of overuse of the default classifier *e* is triggered by changes in the composition of the children's lexicons. In particular, does overuse of *e* begin when a "critical mass" of nouns are learned that require *e* (cf. Marchman and Bates 1994 for inflection)? Again, the answer appears to be no, based on our data. For example, while Lin's first overuse of *e* was recorded approximately the same time as his first recorded correct use of *e* (actually, one month before), he didn't begin his dramatic drop in

classifier accuracy until two months later. This drop doesn't correspond in any obvious way with changes in the proportion of *e* nouns forming his productive vocabulary. Later fluctuations in overuse of *e* show even less correspondence with his vocabulary.

Our findings suggest to us that early classifier acquisition follows a developmental course roughly similar to those found with inflection, including the apparent irrelevance of external factors. Clearly the environment does play some role; for instance, children acquiring Mandarin, whose default classifier is much more commonly used than that in Taiwanese, do not show a U-shaped learning pattern with an initial correct stage, but instead show an S-shaped curve, since they start out using *ge* for virtually all nouns, thus making a large proportion of their earliest productions incorrect (Erbaugh 1986).

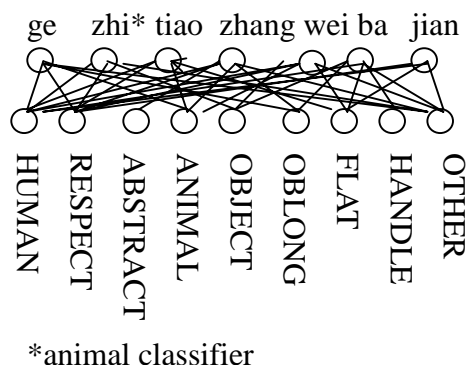
5. Semantic features and connectionist modeling

To better understand the roles of cognition and lexicon in classifier processing and acquisition, we have also been building connectionist models of classifier selection. So far we have just looked at simple two-layer feedforward networks (Plunkett and Elman 1997), where the input vectors represent semantic features for nouns and the output vectors represent different classifiers.

We have been studying the role of extra-linguistic cognition by varying the semantic features used in the networks. In some networks, these features are hand-picked to represent the Mandarin classifier system in a maximally efficient way, the way linguists have described them. For others, we have elicited features from a large set of native Mandarin speakers.

We began by studying models with a very small set of hand-picked features, merely to get a general feel for the problem. In one such "toy" model, we used, as inputs, sets of nine units representing hand-picked binary semantic features and, as outputs, sets of seven units representing common Mandarin classifiers. The network thus had the following overall architecture.

Figure 2. A schematic representation of a two-layer network for classifier selection.



Proportions of each pattern type in the training set were matched with those found in a large corpus of Mandarin text (Chen et al. 1996). Training was done using the **tlearn** software (Plunkett and Elman 1997), which uses the backpropagation learning algorithm. That is, the weights of connections are initially set randomly, so the model initially produces meaningless output for a given input. This is compared against the target output for that input and error is spread back through the network so that weights can be corrected

accordingly. In this way our model gradually learned to associate given combinations of semantic features with particular classifiers.

After training the network, we found the following basic results. First, when two semantic categories were in a subset relation (e.g. "oblong animals" and "animals"), the network used the more specific classifier in the more specific case. Second, when there were competing classifiers for the same input vector, the network consistently chose the classifier that went with that input more often in the training set (e.g. "respected human" triggered output of *wei*, not *ge*). Third, the network generalized the use of the "default" classifier *ge*, as in the case of semantically vacuous nouns represented by a totally empty input. The key to this result is that the *ge* output was the most common output overall in the training set, so that the network was conditioned always to activate the *ge* output node unless other specific input nodes were active.

How do these results match up with what we know about how real people select classifiers? First, real people do indeed show the "subset" effect, but not as strictly as the network. For example, as noted earlier, even with "respected humans" there is a tendency to reduce to *ge*. Second, real people don't choose just one classifier for a noun. In one of the experiments described in Myers, Gong, and Sheng (1999), participants had to choose classifiers for meaningless fake nouns; they chose *ge* 65%-70% of the time when no semantic information was provided, but they also chose others. This probabilistic behavior can be modeled, however, by treating the various output node activations as probabilities rather than looking only at the one node whose activation achieved a threshold. Third, the model did seem able to treat *ge* as a default. This was shown not only by the activation of the *ge* output node for the semantically vacuous input, but also in the fact that MI and activation of the *ge* node were inversely correlated; that is, the model tended to neutralize to *ge* more often for nouns that didn't collocate often enough with an alternative classifier in training. However, in at least one way the model behaved very much unlike real people: as noted above, for inputs containing the feature "respected people", the network's output was *wei*, but for real people the default even in this class continues to be *ge*. That is, for the model, low MI between a given noun and classifier caused it to choose a competitor classifier with higher MI with that noun's features, but for our real experimental participants, low MI caused them fall back on some sort of default *ge*-selection rule.

After achieving a basic understanding with such simple models (especially of the behavior of the default classifier), we followed this up with models that use much more realistic semantic features. We began by using Chinese translations of the English nouns used in the English semantic feature study of McRae, de Sa and Seidenberg (1997), which allowed us to determine which aspects of our features are language-specific and which are presumably extra-linguistic. We thus asked 300 native speakers to give us features, which gave us approximately 150 "raw features" for each of the approximately 200 words. These then had to be pared down and made consistent across nouns in as nonarbitrary a way as possible. Since they were asked to list any properties that come to mind for the nouns they are given, the features were usually much more concrete than those that linguists would propose (e.g. one feature for "airplane" was "crashes").

We have made three general findings from these feature data. First, the Chinese participants almost always named the features that linguists say are necessary for classifier selection. Second, there was a slight Whorfian effect, where the speakers' knowledge of the classifier system seemed to influence their choice of features. As illustrated in the following

table, our Chinese judges occasionally named features that are relevant for classifier choice, but which were not named by McRae et al.'s (1997) American speakers.

Table 2: Whorfian effects in Chinese speakers' feature analyses

Noun	Classifier	Classifier features	Named by Chinese	Named by Americans
dolphin	<i>zhi</i>	animal	yes	no
	<i>tiao</i>	oblong	yes	
corn	<i>gen</i>	oblong	yes	

Third, features relevant for Chinese classifier selection were almost always also named by the American speakers. This implies that these features are truly universally salient, and also makes it plausible that a network trained with these features could learn to imitate human behavior in classifier selection.

However, for our modeling work, one problem with the McRae-based Chinese set is that it contains equal-sized semantic categories. To study *ge* overgeneralization, we also needed a set proportional to real frequencies. Hence we created a new set by selecting pseudorandomly from a frequency-ordered list so as to include as many of the original words as possible; after various nouns were rejected (e.g. synonyms, too unfamiliar, not used with classifiers), we ended up with approximately 500 nouns. In order to avoid the time-consuming and expensive process of analyzing data from another 300 native speakers, we combined features extracted from dictionary definitions and judgments from a mere 30 speakers to give us features (30 words per participant) to supplement our own intuitions. This so-called "proportional" set was used to model apparent default classifier effects, including in acquisition; the "nonproportional" (McRae-derived) set was used to model overall patterns of classifier selection.

In both cases, the network had the same structure as the "toy" model described above: a two-layer network in which the set of input nodes represented features (as chosen by our native speakers), and the output nodes represented classifiers (as chosen by a different set of native speakers). This time, however, the sets of semantic features were much larger; in the nonproportional set (which contained primarily small objects) there were approximately 1200 features, and in the proportional set (which was semantically far more diverse) there were approximately 3500 features. Training again was done using the backpropagation algorithm run in **tlearn**.

To study the behavior of the model with the nonproportional set, we checked its ability to respond correctly to items that it had not been trained on. Using a common protocol in computational linguistics, we trained the model on 90% of the items in the training set and then tested it on the remaining 10%, doing this ten times to cover the entire set of items. The resulting accuracy rate (i.e. selecting the classifier that we required for items it had not been trained on) was a mere 35%, but this was far above the rate expected if the model were merely guessing at random. Moreover, many of its "errors" were completely acceptable alternatives that simply hadn't been in the training set (for example, the "prescriptively correct" classifier for *gongniu* "bull" that we chose was *tou* "head", but the model chose the perfectly acceptable animal classifier *zhi* instead). The model also occasionally chose more than one (correct) classifier at a time (i.e. the activation of more than one output node achieved threshold), again mimicking the probabilistic behavior of real people. In fact, the major problem with the

model was that often no node achieved threshold at all, so that it in effect didn't choose any classifier. This was probably due more to our unrealistically limited set of semantic features (and relative lack of overlap in features across words) than to any inherent limitations in the use of connectionism for this problem.

Misclassification errors made by the model tended to follow semantics, often (as just noted) producing output that was in fact correct. Other times, however, the outputs were unambiguously wrong, such as when it chose the animal classifier *zhi* for *shuzhuo* "desk" instead of the correct "flat" classifier *zhang*. This particular error was probably due partly to the fact that *zhi* happened to have been almost as common in the training set as *ge* (in reality *zhi* is indeed quite common, but not this common), and partly to the peculiar fact that the native Mandarin speakers who gave us the features used in the model did not choose "flatness" as a feature when describing *shuzhuo* (though the Americans did, in a curious inverse of the Whorfian effect).

Surprisingly, given the nonproportional nature of this set (*ge* nouns accounted for fewer than 20% of all items in the training set, far below what is found in actual Mandarin discourse), we still found some evidence indicating a special role for *ge*. For example, as with real people, this model showed significantly more non-*ge* nouns that were neutralized to *ge* than *ge* nouns that received non-*ge* outputs. This pattern was directly related to the slight plurality that *ge* nouns enjoyed in the training set, since there was a significant correlation between the commonness of a classifier in the training set and the likelihood that the model chose this classifier (erroneously) as an output. However, there was no significant difference between the number of errors made with *ge* nouns vs. non-*ge* nouns, and the model also did not show the default generalization effect; when tested on semantically vacuous input (i.e. without specifying any semantic features), the model made no choice of classifier at all; activations for "small object" *ke*, animal *zhi* and *ge* were the highest among all the classifiers (these were also the three most common classifiers in the training set), but none reached threshold.

The proportional set was quite a bit harder to analyze, due not only to the larger number of features, but also difficulties relating to word frequency. In order for frequency to be reflected in the training set, items that were high frequency in the source corpus were repeated more often in the training set than items that were of low frequency in the source corpus. For technical reasons, the particular way in which we did this depended on our ultimate purposes (i.e. modeling adult classifier selection or child language acquisition). For the adult model, we trained the high-frequency words first, and repeated the training multiple times before we input the mid-frequency and low-frequency words. For the child model, this method would have made it impossible to examine change in model behavior over time in a realistic way, so instead we created a single large file containing all of the items, including the frequency-proportional repeats, and trained the model by selecting randomly from this file until the entire file was exhausted, testing the model's behavior at regular intervals during this process.

After training the adult model, we tested it using the method described earlier, where the model was first trained on 90% of the items and then tested on the remaining 10%. Semantically-based misclassification errors were again found, and again many technically "incorrect" outputs were in fact entirely acceptable (e.g. our prescriptive classifier for *shuomingshu* "instruction manual" was *fen* "piece", but the model reasonably chose the "book" classifier *ben*). We also found evidence of the default nature of *ge*, naturally much

stronger than with the nonproportional model (in the proportional model, *ge* formed approximately 30% of all classifier tokens, virtually identical to the written corpus from which the items were taken). Thus as with the nonproportional set, misclassification of nouns was significantly more likely to involve replacement of a specific classifier by *ge* than the other way around. Moreover, this time when a semantically vacuous input was used, the model decisively chose *ge* as the output, just like the toy model. Nevertheless, as with the nonproportional set, there was no significant difference in the error rate for *ge* nouns vs. non-*ge* nouns.

We also found clear frequency effects on accuracy. For high-frequency items, the model was completely accurate for over 40% of the items (again, its accuracy was actually higher when we included acceptable alternatives and simultaneous selections of multiple classifiers), but for lower-frequency items the accuracy dropped down to 25% or so. However, there was no word frequency effect in the likelihood of neutralization to *ge*; this neutralization occurred at approximately the same rate regardless of item frequency, since presumably the key factor for neutralization is classifier-noun collocation, not word-frequency by itself.

Finally we turn to our modeling of classifier acquisition. Since our semantic features, classifier preferences, and proportions all came from Mandarin, we expect that as a network is gradually trained over time, it should mimic the behavior of a child acquiring Mandarin and thus show an S-shaped curve (i.e. an initial stage of overuse of *ge*, followed by gradual improvement), rather than the U-shaped curve discussed earlier with regard to Taiwanese classifier acquisition. This is indeed what we have found with our models. However, such a result is not inherently very satisfying, since connectionist models in general have the tendency to show just such a curve during the acquisition of any data set. In fact, connectionist models of U-shaped development in English inflection (e.g. Plunkett and Marchman 1993) have, in essence, been accused of cheating in order to obtain an initial stage of correct performance (e.g. Marcus 1995), namely by modifying the training set over time to allow the model to get a good grasp on a subset of the items before being presented with new items. This method is considered legitimate by its practitioners since it is intended to model the child's changing "uptake" of activate vocabulary rather than the steady, unchanging input from the parents, but it remains highly controversial. Moreover, at most such models only show what is called "micro-U-shaped" development, where accuracy drops off slightly for a single word or subset of words and immediately rises again.

Putting our study of classifier acquisition into this context, it would be relatively straightforward for us to attempt a model of Taiwanese classifier acquisition, assuming the same features used in Mandarin, but training with different classifiers in different proportions. It seems unlikely that the empirically attested U-shape would emerge naturally any more than it has in models of English inflection, but this is still an open question.

6. Concluding remarks

In this brief paper we did not intend to describe all of our work on Sinitic classifier systems in exhaustive detail, but rather merely sketch out the range of our work to date, some of which is already publicly available, and some of which is still work in progress. In any case, we hope that this quick survey has convinced at least some readers that research in this area has the potential to shed light on many deep issues in the study of language, mind, and the relationship between them.

References

- Ahrens, K. 1994. "Classifier Production in Normals and Aphasics." *Journal of Chinese Linguistics*. 22.203-247.
- Ahrens, K., and C-R. Huang. 1996. "Classifiers and Semantic Type Coercion: Motivating a New Classification of Classifiers." *Language, Information and Computation (PACLIC 11)*. 1-10.
- Aikhenvald, A. Y. 2000. *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.
- Allan, K. 1977. "Classifiers." *Language*. 53.285-311.
- Bates, E. and J. C. Goodman. 1997. "On the Inseparability of Grammar and the Lexicon: Evidence from Acquisition, Aphasia and Real-time Processing." *Language and Cognitive Processes*. 12.5-7-584.
- Carpenter, K. 1991. "Later Rather Than Sooner: Extralinguistic Categories in the Acquisition of Thai Classifiers." *Journal of Child Language*. 18.93-113.
- Chang-Smith, M. 2000. "Empirical Evidence for Prototypes in Linguistic Categorization Revealed in Mandarin Numeral Classifiers." *Journal of Chinese Language Teachers Association*. 35.19-52.
- Chao, Y. R. 1968. *A Grammar of Spoken Chinese*. Berkeley, CA: University of California Press.
- Chen K-J, C-R Huang, L-P Chang, and H-L Hsu. 1996. "SINICA CORPUS: Design Methodology for Balanced Corpora." *Language, Information and Computation*. 11.167-176.
- Cheng, L. L.-S. and R. Sybesma. 1998. "Yi-wan Tang, yi-ge Tang: Classifiers and Massifiers." *Tsing Hua Journal of Chinese Studies, New Series*. 28.385-412.
- Church, K. W. and P. Hanks. 1990. "Word Association Norms, Mutual Information, and Lexicography." *Computational Linguistics*. 16.22-29.
- Erbaugh, M. 1986. "Taking Stock: The Development of Chinese Noun Classifiers Historically and in Young Children." In C. Craig [ed] *Noun Classes and Categorization*, 399-436. John Benjamins.
- Erbaugh, M. 2000. "Classifiers Are for Specification: Complementary Functions for Sortal and General Classifiers in Cantonese and Mandarin." *Proceedings of the 33rd meeting of the International Conference on Sino-Tibetan Languages and Linguistics, Bangkok, October 2-6*.
- Hu, Q. 1993. *The Acquisition of Chinese Classifiers by Young Mandarin-speaking Children*. Boston, MA: Boston University Ph.D. Dissertation.
- Lakoff, G. 1986. "Classifiers as a Reflection of Mind." In C. Craig [ed] *Noun Classes and Categorization*, 13-51. John Benjamins.
- Landau, B. 1996. "Multiple Geometric Representations of Objects in Languages and Language Learners." In P. Bloom, M. A. Peterson, L. Nadel, and M. F. Garrett [eds] *Language and Space*, 317-363. Cambridge, MA: MIT Press.
- Langacker, R. W. 1991. *Foundations of Cognitive Grammar*. Stanford University Press.
- Li, C. N. and S. A. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.
- Loke, K.-K. 1994. "Is *Ge* Merely a 'General Classifier'?" *Journal of the Chinese Language Teachers Association*. 29.35-50.
- Marchman, V. A. and E. Bates. 1994. "Continuity in Lexical and Morphological Development: A Test of the Critical Mass Hypothesis." *Journal of Child Language*. 21.331-366.
- Marcus, G. F. 1995. "The Acquisition of the English Past Tense in Children and

- Multilayered Connectionist Networks." *Cognition*. 56.271-279.
- Marcus, G. F., U. Brinkmann, H. Clahsen, R. Wiese, and S. Pinker. 1995. "German Inflection: The Exception That Proves the Rule." *Cognitive Psychology*. 29.189-256.
- Marcus, G. F., S. Pinker, M. Ullman, M. Hollander, T. J. Rosen, and F. Xu. 1992. Overregularization in Language Acquisition. *Monographs of the Society for Research in Child Development* 57.
- McRae, K., V. R. de Sa, and M. S. Seidenberg. 1997. "On the Nature and Scope of Featural Representations of Word Meaning." *Journal of Experimental Psychology: General*. 126.99-130.
- Myers, J. 2000. "Rules vs. Analogy in Mandarin Classifier Selection." *Language and Linguistics*. 1.187-209.
- Myers, J., S-P Gong, and Z-G Shen. 1999. "The Semantic Content of the General Classifier in Mandarin." Presented at Presented at the International Association of Chinese Linguistics Eighth Annual Conference, Melbourne, Australia.
- Myers, J. and J. Tsay. 2000. "The Acquisition of the Default Classifier in Taiwanese." *Proceedings of the International Symposium on Chinese Languages and Linguistics*. 7.87-106.
- Newmeyer, F. J. 1998. *Language Form and Language Function*. Cambridge MA: MIT Press.
- Ng, B. C. 1989. *The Acquisition of Numeral Classifiers in Hokkien*. La Trobe U. (Australia) Ph.D. Dissertation.
- Pinker, S. 1991. "Rules of Language." *Science*. 253.530-535.
- Pinker, S. 1994. *The Language Instinct*. William Morrow & Co.
- Pinker, S. 1999. *Words and Rules: The Ingredients of Language*. Basic Books.
- Plunkett, K. and J. L. Elman. 1997. *Exercises in Rethinking Innateness: A Handbook for Connectionist Simulations*. MIT Press.
- Plunkett, K. and V. Marchman. 1993. "From Rote Learning to System Building." *Cognition*. 38.43-102.
- Polio, C. 1994. "Non-native Speakers' Use of Nominal Classifiers in Mandarin Chinese." *Journal of the Chinese Language Teachers Association*. 29.51-66.
- Prasada, S., and S. Pinker. 1993. "Similarity-based and Rule-based Generalizations in Inflectional Morphology." *Language and Cognitive Processes*. 8.1-56.
- Rumelhart, D. and J. McClelland. 1986. "On Learning the Past Tenses of English Verbs: Implicit Rules or Parallel Distributed Processing?" J. McClelland, D. Rumelhart, and the PDP Research Group [eds] *Parallel Distributed Processing*. Cambridge, MA: MIT Press.
- Sereno, J. A. and A. Jongman. 1997. "Processing of English Inflectional Morphology." *Memory and Cognition*. 25.425-437.
- Tai, J. H-Y. 1992. "Variation in Classifier Systems across Chinese Dialects." *Chinese Languages and Linguistics 1: Chinese Dialects*, 587-608. Academia Sinica.
- Tai, J. H-Y. 1994. "Chinese Classifier Systems and Human Categorization." M. Y. Chen and O. Tzeng [eds] *In Honor of William S.-Y. Wang: Interdisciplinary Studies on Language and Language Change*, 479-494. Taipei: Pyramid Press.
- Tai, J. H-Y. 1997. *Categorization Patterns of Classifiers in Taiwanese Southern Min and Their Cognitive Principles*. [in Chinese] National Science Council project report, Taiwan, NSC-86-2411-H-194-005.
- Tai, J. H-Y. and F.-Y. Chao. 1994. "A Semantic Study of the Classifier *Zhang*." *Journal of Chinese Language Teachers Association*. 29.67-78.
- Tai, J. H-Y. and L. Wang. 1990. "A Semantic Study of the Classifier *Tiao*." *Journal of Chinese Language Teachers Association*. 25.35-56.

- Tien, Y. M. 1996. The Analysis of Semantic Verification Process between Classifiers and Nouns in Mandarin Chinese. [in Chinese]. National Chung Cheng University masters thesis.
- Tsay, J. 1997-2000. A Developmental Study of Tone Acquisition in Taiwanese. National Science Council project, Taiwan. NSC-87-2411-H-194-019, NSC-88-2411-H-194-019.
- Tsay, J. 2001-2003. Taiwan Chinese Language Corpus. National Science Council project, Taiwan. NSC 89-2411-H-194-067, NSC 90-2411-H-194-031.
- Tzeng, O., S. Chen, and D. L. Hung. 1991. "The Classifier Problem in Chinese Aphasia." *Brain and Language*. 41.184-202.
- Ullman, M. T. 1999. "Acceptability Ratings of Regular and Irregular Past-tense Forms: Evidence for a Dual-system Model of Language from Word Frequency and Phonological Neighborhood Effects." *Language and Cognitive Processes*. 14.47-67.
- Wiebusch, T. 1995. "Quantification and Qualification: Two Competing Functions of Numeral Classifiers in the Light of the Radical System of the Chinese Script." *Journal of Chinese Linguistics*. 23.1-41.
- Zubin, D A. and M. Shimojo. 1993. "How 'General' Are General Classifiers? With Special Reference to *ko* and *tsu* in Japanese." *Proceedings of the Berkeley Linguistics Society*. 19.490-502.