FUNCTION FOR SORTAL AND GENERAL CLASSIFIERS IN CANTONESE AND MANDARIN

Erbaugh, Mary S.

ABSTRACT. Classifiers specify a noun as focus of attention. They are not triggered by gender-like rules, for some 40% of nouns appear without any classifier in matched Cantonese and Mandarin narratives. Classifiers add information by appearing with the noun rather than as a pro-form (97% Cantonese, 99% Mandarin). Sortal classifiers, e.g. 'extended object' *tiuh* 條, are rare. Just 18% of Cantonese nouns and 3% of Mandarin appear with a sortal. Sortals mark an event focus on new information, on objects of action verbs (63% Cantonese, 81% Mandarin). The general classifier appears in complementary distribution with existential or stative verbs. Cantonese uses six times more sortals than Mandarin, largely on later mentions of the same object. Though brain imaging can map categories which roughly correspond to classifiers, a frequency-weighted, connectionist analysis provides the best model. **Keywords:** classifiers, discourse, connectionism, Cantonese, Mandarin.

THE FAILURE OF CATEGORICAL RULES FOR CLASSIFIERS [1]

Both prescriptive and theoretical grammars assume that a noun classifier appears with every Chinese noun like gender in Romance languages. Dictionaries list the gender of every noun, for grammatical gender is automatic, obligatory and rule-governed for every noun, article and adjective. [2] But the gender model of categorical rules fails for classifiers (which are overviewed in Appendix 1). It fails to explain why sortal classifiers such as 'extended object-CL' *tiuh* 條 are rare. Some 40% of all nouns appear without either a general or sortal classifier when Cantonese and Mandarin speakers describe the same short 'Pear Stories' film (described in Appendix 2). In fact, 40% of the Mandarin speakers describe the whole film using only the general classifier (as do 3% of Cantonese). Speakers also disagree about even core classifiers. No two Mandarin speakers among 50 people tested agreed about which classifiers to use for the same set of objects (Bourgerie 1996).

This variability contradicts the classical theory of meaning, dominant from Aristotle on, which posits that each word and object possess a single, distinct defining feature (Taylor 1995). Classifiers once seemed a reasonable candidate for the Aristotlean features. Dictionary listing of classifiers encouraged this view. When more sophisticated views of prototype semantics superseded the classical approach, the search moved toward superordinate categories, such as 'animal' or 'vehicle' to subcategorize basic exemplars such as 'dog' or 'car'. However, more detailed examination finds that few classifiers work well as superordinate categories. Most define very heterogeneous, overlapping sets. In addition, the same object can often take any of several classifiers (Matthews and Yip 1994:105-07). And the same speaker may use different classifiers for the same object with no discernible change in viewpoint. In the Pear Stories, the goat, on screen for just seven seconds, is classified in Mandarin with the 'head-CL' *tou* 頭, the 'small animal/small object-CL' *zhi* 隻, as well as the 'extended object-CL' *tiao* 條, used also for the road and a string.

Classifier use is so irregular that sortals are not a reliable test of language damage or development. When researchers elicit classifiers by showing a picture, e.g. three chickens, both children and adults often respond with a bare noun, 'chicken[s]' *gai/ji* 雞. Pressed to answer 'how many', they often reply 'three' with the general classifier *go/ge* 個, avoiding the desired 'three animal-CL chickens'. (Omitting a general classifier where grammatically required *is* a valid measure of normal development.)

In conversation, even highly educated people often use a general classifier where prescriptive grammar requires a sortal. In the Pear Stories, these included the hat, bike, goat, and tree (1% of Cantonese nouns, 3% Mandarin). More important, the general classifier is the only possible choice for the overwhelming majority of nouns, not just abstractions like 'idea' or 'revolution', or large unique entities such as cities, lakes, or planets; but also for a huge

ercentage of small concrete nouns, including pears, leaves, baskets, ladders, wheels, and oys. Even dictionaries of classifiers list only a small subset of nouns. Investigation is ngoing, but perhaps as few as 20% of nouns can even take a classifier.

Yet classifiers remain psychologically salient despite their low frequency. Jokes abound bout dialect speakers, such as the Hakkas, referring to humans with the classifier which antonese and Mandarin speakers use for animals and small objects *jek / zhi* 隻. Classifiers re often named as speech errors typical of children, substandard speakers, and foreign tudents. Classifiers are salient not just because of grammatical theory, but because they eflect the noun superiority effect, the term psychologists use for the universal tendency to emember concrete nouns better, name them faster as examples, and use them in metaphors.

CLASSIFIERS ARE FOR SPECIFICATION. Why do the categorical rules which work o well for gender fail for classifiers? The puzzle is solved when we discard the seductive otion that classifiers exist primarily to categorize objects, and examine their discourse unctions. A complementary system emerges between sortal and general classifiers, which ppear only when a noun is specified as a focus of individuated attention. Such focus is ecessarily variable, not categorical. Cantonese and Mandarin display a striking common attern, as Table 1 indicates. About 40% of nouns appear without a classifier (39% Cantonese, 43% Mandarin). And so, virtually identical proportions appear with a classifier 62% Cantonese, 57% Mandarin). The general classifier appears with 44% of Cantonese ouns, 54% of Mandarin. There is no statistically significant difference between these roportions.

TABLE 1 COMPLEMENTARY CLASSIFIER SYSTEM

	NOUNS with NO CLASSIFIER	GENERAL 個	SORTAL (e.g. 條)
CANTONESE	39%	44%	18%
MANDARIN	43%	54%	3%

Sortal classifiers are used with nouns for new information, typically as objects of an ction verb. General classifiers, in contrast, specify an existential, translatable as 'there's a ....' hey appear disproportionately as subjects of stative verbs. This striking commonality is liscussed further below, as is the much greater use of sortals in Cantonese. Neither Cantonese or Mandarin speakers ever omit a classifier where grammatically required after a numeral or a leterminer. But a bare noun is often correct, especially for generic or non-specific reference uch as 'I'm busy writing a report' / 'I'm busy report-writing', Mandarin 'I busy-PROGRESSIVE vrite report' *wo mangzhe xie baogao* 我 忙著 寫 報告(Chen 1987).

DIFFERENCES BETWEEN CANTONESE AND MANDARIN CLASSIFIERS

Cantonese and Mandarin classifiers differ far more than previously assumed. Semantic lifferences have received the bulk of attention. The Cantonese for 'needle' *jam* 針 takes 'eye-CL' *ngaan* 眼 . A Mandarin needle, written with the same character but pronounced *zhen,* akes 'root/thread-CL' *gen* 根 . Grammatical differences also exist. Cantonese classifiers can e used alone, without a determiner (Matthews and Yip 1994:92-109, Shi 1996, Zeng 1982, hou 1997). Cantonese for 'that book cannot be found' is, 'volume-CL book not perceive ERFECTIVE-ASPECT' *bun syu m gin jo* 本 書 唔 見 咗. (Mandarin: That volume-CL book not erceive PERFECTIVE-ASPECT' *na ben shu bu jianle* 那 本 書 不 見 了 .)

Most sortals refer to more heterogeneous sets than 'book' or 'hat'. Brief glosses only int at the complexity. Cantonese sortals refer to more mixed groups than Mandarin, in part ecause Cantonese is unstandardized and seldom written. The Mandarin *zhi* 隻 classifies irds, and other small, animals such as rabbits and cats. Tigers and leopards are included, by xtension from cats; in addition to deer. Ears, hands, shoes and socks, also take *zhi,*

probably by extension from smallness and manipulability. In Cantonese, the same character, pronounced *jek*, covers both small and large animals, including horses and oxen, as well as teeth, eggs, battleships, and phonograph records. In other cases the same character is used as classifier for a different but related meaning. Mandarin *jian* 間 classifies individual rooms. In Cantonese, the same character, pronounced *gaan*, refers to a whole flat or building. Even the semantics of the general classifier differ. Cantonese *go* 個 is strongly associated with verticality, and humans, who are 90% of its referents in the Pear Stories. In Mandarin, only 64% of the general classifier *ge* 個 has human referents (difference significant at $p < .01$).

GREATER USE OF SORTALS IN CANTONESE. Cantonese speakers unexpectedly use six times more sortal classifiers per noun, a highly significant difference ($p < .001$). [3] Table 2 summarizes the contrast, while Tables 3 and 4 offer details for all speakers. The 19 Mandarin narratives yielded only 35 sortals, just two per narrative on average, or one every 30 nouns. [4] The 30 Cantonese narratives yielded 241 sortals, 8 per narrative, one every six nouns. Just 5% of the Mandarin classifiers were sortal, in contrast to 28% of the Cantonese. The difference is robust, because it emerged even though the Hong Kong women told much shorter stories. [5] And Mandarin, not Cantonese, serves as the high register, close to prescriptive written style. However, the greater Cantonese usage of sortals comes largely on later mentions of the same object.

TABLE 2 SORTAL VERSUS GENERAL CLASSIFIERS

	CANTONESE	MANDARIN
Sortals -- mean per narrative	**8**	**2**
Percentage of classifiers general	**72%**	**95%**
Percentage of classifiers sortal	**28%**	**5%**
Ratio of general to sortal classifiers	**2:1**	**20:1**

GREATER RANGE OF SORTALS. The Cantonese also use a 15% greater range of sortals, 14 compared to 12 in Mandarin. Five (35%) are distinctively Cantonese: 'tree-vegetable-CL' for the pear trees *po* 合; 'lump-CL' *gaau*㘝舊 for the rock and a ball; 'performance-CL' *cheut* 齣 and 'set-CL' *tou* 套 for the film; and 'burst-CL' johk 濯 for a breeze. Cantonese classifier dictionaries show even greater difference. Some 62% of the sortals in the most comprehensive classifier dictionary do not exist in Mandarin, or differ substantially (Killingley 1982); 68% are differ in another source (Zeng 1982).

Cantonese also has a greater overall range of sortals in common use. Mandarin, closer to the written language, is the favorite of dictionary makers. One classifier dictionary lists 199 classifiers, 76 of them sortal (38%) (Wang and Wu 1989). A teaching grammar of 'the 143 most commonly used classifiers' lists 66 sortals (46%) (Jiao 1993). Cantonese has well over 80 sortals. One guide lists 87 classifiers, 59 sortal (69%) (Zeng 1982). The most extensive classifier dictionary lists 157 classifiers, 102 of them (65%) sortal, or a combination of sortal and measure (Killingley 1982). Cantonese appears to have a greater range of classifiers for transient events such as smells, sounds, and flashes of light.

Size and shape classifiers have attracted much attention, especially for Mandarin. They form 27% of classifiers in the Pear Stories, appearing on just under 1% of nouns (*tiao* 條, *kuai* 塊, *duan* 段, *pian* 片). [6] Cantonese size and shape classifiers comprise 14% of sortals (*gaau* 㘝舊, *tiuh* 條, *faai* 塊, *dyuhn* 段). This lower percentage merely reflects the greater range of classifiers. Size and shape is at least as important in Cantonese, for they appear on 2% of nouns.

Cantonese and Mandarin speakers mention almost exactly the same objects and events. But they differ in which objects take a sortal, and which sortals they choose (Table 5). The five most frequent Cantonese classifiers account for 81% of all tokens. In descending

rder, they refer to the hat, the bike, the tree, the goat, and the rock. ('Hat/peak-CL' *ding* 頂, nachine/frame-CL' *ga* 架, 'tree-CL' *po* 㘝, 'animal-small thing-CL' *jek* 隻, 'lump-CL' *gaau* 舊.) The Mandarin top five are less concentrated, 62% of tokens. They refer to the bike, the ree, the sound of the whistle, the goat, and the road. (The bike takes 'vehicle-CL' *liang* 輛, he tree takes 'tree CL' *ke* 棵, the whistle takes 'sound-CL' *sheng* 聲, the goat takes 'head-CL' *ou* 頭, 'animal or small thing-CL' *zhi* 隻, and the 'extended object-CL' *tiao* 條. The road is qually likely to take the 'extended object-CL' or the 'section-CL' *duan* 段.) Much previous esearch considers the culturally valued objects which are more likely to receive classifiers: orn in Mayan, elephants in Thai. But the highest frequency classifiers are overwhelmingly anal. A large sample of adult Mandarin conversation found that speakers used only 22 ifferent classifiers (Erbaugh 1986).[7]

SHARED PATTERNS OF INFORMATION DEPLOYMENT

IDEA UNITS AND SPECIFICATION OF NEW ENTITIES. Universals of information rocessing constrain both Cantonese and Mandarin. Speech emerges in short bursts of five or ix words, which Chafe (1980) calls idea units, followed by a pause and/or falling intonation. he length, organization of idea units, and pause patterns of the Mandarin Pear Stories trikingly match the English stories, with Chinese speakers varying even less than the mericans (Cumming 1984, Erbaugh 1990). Narrators begin with an existential phrase such s 'There's a....' 'to have/to exist one + CL' 有一個. Though previous studies treated such hrases as indefinite, Matthews and Pacioni demonstrate that the classifier picks out a specific tem (1997). 'One' functions not as a numeral, but as a specifier for the main character. The epetition of the tree classifier specifies the tree, first as a location of action, then as the armer's property:

(1a) 我，唉。
ngoh, eih
I, uh.

(1b) 睇 到 有 一個 。。。 一個 男人 啦。
tai dou yauh yat go .... yat go naahmyahn la
See reach exist one gen-CL... one gen-CL male person PARTICLE.

(1c) 咁 佢 應該 係 一 㘝 樹，
gam keuih yinggoi hai yat po syuh,
Well, he must be in one tree-CL tree,

一 㘝 梨 樹 嘅 主人 類 嘅。
one TREE-CL pear tree POSS owner sort POSS.
'I, uh. [I] see there's a ... um ... a man. Well, he must be in this tree, this pear tree's owner, that sort of thing'.

This initial framing is effortful, full of hesitations and substitutions. Later narration lows more smoothly, dividing into episode boundaries which closely match film cuts to ifferent scenes. Classifiers tend to reappear at episode boundaries to re-establish reference.

Adults, preschool children and even foreign students follow the same principle, which is ever explicitly taught, of choosing a sortal or general classifier based on the hearer's nformation needs (Erbaugh 1986, Polio 1994). A sortal is especially like to appear with a noun or a new topic, especially for objects which are unfamiliar to the hearer or not physically resent. In Mandarin, a bike viewed out the window might be called 'that vehicle-CL bike' *na iang zixing che* 那輛 自行 車 But a bike parked in the living room, or a tiny toy bike on the