REDUPLICATION IN MANDARIN CHINESE:

THEIR FORMATION RULES, SYNTACTIC BEHAVIOR

AND ICG REPRESENTATION

Feng-yi Chen*, Ruo-ping Jean Mo*, Chu-Ren Huang**, Keh-Jiann Chen*

*The Institute of Information Science, Academia Sinica
**The Institute of History and Philology, Academia Sinica
Nankang, Taipei, Taiwan,
Republic of China

ABSTRACT

Morphologically derived words can neither be identified by dictionary look-up nor be accounted for with a syntactic parser in NLP. Mandarin Chinese involves several productive morphological rules. This paper proposes a set of rules to identify reduplicatives in Mandarin Chinese. This set of rules will be used to complement dictionary look-up and DM generation rules in the word segmentation module. The co-occurrence restriction of adjuncts in reduplication is also discussed and expressed in ICG mechanism to improve parsing results.

I. Introduction

Mandarin Chinese reduplicatives are constructed by repeating the whole or part of a lexical item. Verbal reduplicatives may denote delimitative as well as tentative aspects(嚐 試 貌) or intensifying meaning. For instance, both verbs *chang* 'to sing' and *pingan* 'to be safe and secure' can be reduplicated, as shown in (1).

> (1) a. tamen pingshr shihuan *chang chang ge*
>   they usually like   sing sing song
>   'They usually like to sing a little bit'
>
>   b. shiauhai dou *ping ping an*   *an*
>   children all flat   flat peaceful peaceful
>   'Children are all very safe and secure'

In addition to verbs, onomatopoeia, measure words, and morphological derived determinative-measure compounds can also undergo the process of reduplication. Because of its high productivity and its being fed by another morphological rules, exhaustively listing reduplicatives in the lexicon is not a viable alternative.

The current version of CKIP word segmentation system [3] is based on a lexicon of about ninety thousand words and a set of determinative-measure rules [7]. Without rules to account for reduplicatives, not all correct word breaks can be found. Example extracted from machine-readable Chinese corpus is given in (2)[1].

> (2) a. juang man le *yi dai dai* de pingguo
>   fill   full LE one bag bag DE apple
>   '(The container etc.) is filled with bags of apples.'

---

1. The following examples presented in this paper are mainly adopted from this machine-readable corpus whose texts are mostly from a Chinese newspaper, *Tz You Shr Bau* 'Liberty Times' from October 1990 to February 1991, which contains 10 million words or so.

b. <u>yijr</u>   <u>tzai</u> gung <u>shiu</u>   <u>shrchang</u> jung   *jang jang* <u>*die*</u> <u>*die*</u>
always in supply demand market center rise rise fall fall
'(The price) has always been <u>rising and falling alternatively</u> in the market.'

In order to solve the above-mentioned problem, this paper proposes a set of reduplicative-formation rules to build reduplicatives within the word segmentation module. We will first discuss the scope and types of reduplicatives in Mandarin Chinese, and their syntactic behavior and semantic variations.

II. Reduplicatives - their scope and types

Morphologically, reduplicatives are formed by total or partial repetition of a lexical item. However, not every lexical item containing partial repetition is regarded as a reduplicative in this paper. This is because reduplicatives are handled in terms of morphological rules in this paper. And some of the reduplicated types do not follow these requirements: they have limited productivity and show idiosyncratic grammatical behavior, which mean their forms are not predictable by general rules (cf. 3a, 3b, 3c, and 3d).

(3) a. the reduplication of adverbs

chang chang
often often
'often'

jin jin
only only
'only'

b. the reduplication of nouns

en   en   yuan yuan
favor favor hatred hatred
'gratitude and grudge'

219

> *shr   shr   wu   wu*
> thing thing object object
> 'things and objects'

c. the xlixy reduplicatives

> *luo   li   luosuo*
> chatter inside wordy
> 'verbose or wordy'

> *tu   li   tuchi*
> earth inside rustic
> 'rustic'

d. the xyy reduplicatives

> *liu   you you*
> green oil oil
> 'bright green'

> *shiau ha ha*
> laugh Ha Ha
> 'laugh heartily'

> *leng bing bing*
> cold ice ice
> 'icy'

> *shie   lin   lin*
> blood drench drench
> 'bloody'

In (3c), for example, only those adjectives with pejorative meaning, such as *tuchi* 'rustic' and *luosuo* 'wordy', can undergo this type of reduplication. Since their set is quite small, we will simply list them all in the lexicon. As for the xyy type, the meaning as well as the reduplicated yy form are lexically determined by the head x, which may be a noun, verb or adjective [1]&[4]. Try to compare the examples below.

(4) a. *hei   chi   chi*
    black paint paint
    'very dark'

220

*a'. *bai* <u>*chi chi*</u>
   white paint paint

b. *ching* <u>*piau piau*</u>
   light float float
   'lightly'

*b'. *jung* <u>*piau   piau*</u>
   heavy   float float

From example (4), we may observe that the reduplicated yy types, <u>*chi chi*</u> and <u>*piau piau*</u> must co-occur with *hei* 'black' and *ching* 'light', respectively. Using regular expressions to construct these reduplicatives is not feasible because there is no context-free constraint to rule out the non-existing forms. Therefore, they will still be stored in the lexicon.

To sum up, only those reduplicatives of high productivity and predictability are formed by rules and are included in this paper. Moreover, the categories that can have reduplication are limited to verbs, determinative-measure compounds, measures and onomatopoeia.[2] In what follows, a detailed discussion of various reduplicated forms and a set of formation rules will be offered according to different meaning properties.

2.1. Reduplication to Express Tentative Aspect

Generally speaking, the process of reduplication may add a sense of tentativeness to any action verbs which contain no modifier-head internal structure and have no meaning contradicted to the semantic function of reduplication, such as controllable verbs. Lastly, a reduplicatable action verb

---

2. Among these four categories that can undergo reduplication, except for the verb's, the forms of the other three are rather simple. Hence, in the paper, we will only list their formation rules without further explanation. As for the semantic functions of these reduplicatives, a sense of vividness is imposed if the original form is an onomatopoeia; otherwise, 'each' or the way of measuring is added.

cannot be an achievement verb. Take *chujia* 'become a monk' for example. Although this verb is both active and controllable, it cannot be reduplicated because it is an achievement verb. This verb is an achievement verb because the goal of becoming a monk is attained at the endpoint of the denoted action. In terms of lexical semantics, a tentative aspect stipulates that an act be broken down and carried out in a piecemeal fashion. The instantaneity of an achievement verb contradicts this interpretation. This is how our rule excludes *chujia* 'become a monk'. Additional functional types as shown in the following sections.

2.1.1. XX Type[3]

If the input is a monosyllabic action verb, some other words like *yi* 'one', or *le* 'PERFECTIVE' may be inserted between the two Xs; but if the original form is a disyllabic one, then no word may occur in between. Examples are shown in (5).

> (5) a. *shiou (yi) shiou* ye hua de shiangwei
>   smell (one) smell wild flower DE fragrance
>   'try and smell the fragrance of wild flowers a little'
>
>   b. *jengli jengli* shuguei
>   arrange arrange bookcase
>   'arrange the bookcase a little'

Though there is a little difference in forms and number of syllables, only one rule is proposed.

> RD1 --> X ({*yi, le*}) X
>   conditions: (1) X = VA, VB, VC, VD, VE, VF
>   (Action Verb)

---

3. The capitalized X used here means that X may be a syllable or a word and x, y or z just represents a syllable.

222

(2) IF X is monosyllabic
THEN {*yi, le*} is allowed

However, from the data collected, we find that some disyllabic stative verbs may also take this XX form. Closer investigation shows that their reduplicated counterparts are more active-like. The interpretation of the reduplicated (6c) is derived from the active reading of *Kelian* "to pity" in (6b), not the stative reading in (6a). Thus the generalization that only action verbs have tentative reduplicative forms is correct.

(6) a. ta shiangdang *kelian*
    he quite   pitiful
    'He is quite pitiful.'

   b. jingfang shiangdang *kelian* dueifang
      police quite   pitiful other-side
      'The police pity the other party very much.'

   c. *kelian kelian* women
      pitiful pitiful we
      'Just pity us a little bit!'

2.1.2. xxy(z) Type

Only VO compounds are allowed to take this reduplicated type. After reduplication, the result also has an additional meaning of 'doing a little bit', just like the tentative aspect reduplication of XX.

(7) a. ching ta lai   *ping ping li*
    please he come judge judge reason
    'Just ask him to come (and try) to make a judgement.'

   b. shiuang minjung lai   *kai kai yanjie*
      hope   populace come open open view
      '(We) wish the people will come and have their perspectives
        (somewhat) widened.'

These VO compounds may be two or three syllables, so the reduplicated type may either xxy or xxyz. And the formation rule is expressed in the following.

RD5 --> x ({*le, yi*}) xy(z)
    conditions: xy(z) = VA13, VA3, VA4, VB

2.2. Reduplication to Express Vividness

The reduplicated process will impose a sense of vividness to stative verbs, or to intensify the attributes described by them. The major types of reduplicated stative verbs are presented below.

2.2.1. xx Type

Monosyllabic stative verbs can have this reduplication form. However, unlike monosyllabic action verbs, only *de* 'DE' or *di* '-ly' are allowed to follow it, as in (8). To distinguish reduplicated stative from action verbs, RD2 is proposed.

(8) dutz *kung kung (de)*
    belly empty empty
    'The belly is empty.'

RD2 --> xx ({*de, di*})
    conditions: x = VH (Stative Verb)

2.2.2. xxyy Type

In general, only disyllabic stative verbs may undergo this kind of reduplication, and except for few which take sentential complements, most of

224

them fall under the category of intransitives.

(9) a. dajia   dou hen *kaishin*
       everybody all very happy
       'Everybody is very happy.'

   a'. dajia   dou *kai kai shin shin*
        everybody all ha- ha- ppy- ppy
        'Everybody is very happy.'

   b. ta hen *gaushing* toutzren dou neng dacheng gungshr
       he very glad   investor all able   attain   concerns
       'He is very glad that investors can reach concerns.'

   b'. rang dajia   dou neng *gau gau shing   shing* (de)
        let everybody all able ha- ha- ppy-   ppy
        'Let everybody be very happy.'

The rule to form this type of reduplication is expressed in RD7.

RD7 --> xxyy
        conditions: xy = VH11, VH21, DH[+onomatopoeic]

But, after examining more linguistic data, we discovered that some of the
action verbs, whether intransitive or not, can also have this xxyy reduplicated
type. They spread sporadically over the whole active set.

(10) a. yin   juntz   *jin   jin   chu chu* diauyu chang
         addiction gentleman enter enter out out fish   field
         '(Addicted) smokers frequent fishing arenas.'

    b. *yau   yau   huang   huang* bu   chu fangjian
        shake shake wobble   wobble walk out room
        '(S/he) walked out the room unsteadily.'

The semantic function of this reduplicated type also differs with respect to the
input: if the original word is stative, then the morphological process will make

it sound more vivid. Besides, the attributes it describes may also be intensified. But to those active input, only a sense of "doing a little bit' is added.

2.3. Reduplication of Onomatopoeic and Measure Words

Onomatopoeic words and measure words can also have XX reduplicated form. They are accounted for with RD4 and RD10, respectively. The Kleene stars '*' in rules means that onomatopoeic or measure words can be repeated twice or more to form reduplication. RD4 can be applied to both monosyllabic and disyllabic onomatopoeia.

> RD4 --> DH* ({*de, di*})
> conditions: DH (manner adverb) which is specified with the
> feature [+onomatopoeic]

> RD10 --> RNOP1 ((*you*)) RNOP1)*

> RNOP1 --> (IN1)(DESC) M⁴

It deserves mentioning that disyllabic onomatopoeia can undergo reduplication of xxyy type. The input to xxyy-type onomatopoeic reduplication has already been accounted for in RD7.

III. The Syntactic Constraints of Reduplication

Unlike typical morphological processes, Mandarin Chinese reduplication does not change either the argument structure or the category. This rule changes the semantics and some minor syntactic behavior, such as the allowed adjuncts and syntactic patterns. This section concerns with these reduplicated constructions and their representation in the Information-based

---

4. According to Mo et al. [7] 'IN1' in this rule represents numeral compounds and 'DESC' are descriptive words, such as *da* 'big', or *shiau* 'small'.

Case Grammar (ICG), which is proposed by Chen and Huang [2].

3.1. Distribution of Reduplication

3.1.1. Reduplicated Stative Verbs

Though both reduplicated and unreduplicated forms can function as main predicates and manner adverbs, the usage of reduplicated stative verbs is more restricted than their counterparts. Words after being reduplicated can neither co-occur with degree adverbs nor appear in the construction of comparison. This suggests that reduplication assigns the semantic feature of [-SCALE] because non-scaler predicates can neither occur in a comparative construction nor be modified by a degree adjunct [5]. In other word, reduplication turns a scaler predicate into an absolute predicate.

   (11) a. yanjing *shiau shiau* de
         eye   small small DE
         '(His/her) eyes are very small.'

    *a'. yanjing <u>feichang</u> *shiau shiau* de
         eye   very   small small DE

    b. tamen fuchi liang yishiang *ping ping shuen shuen* de
        they couple both always flat flat smooth smooth DE
        'They have always been going smoothly as a couple.'

    *b'. tamen fuchi liang yishiang <u>hen</u> *ping ping shuen shuen* de
        they couple both always very flat flat smooth smooth DE

    c. wo *gau* ta san gungfen
      I tall he three centimeter
      'I am three centimeter taller than he'

    *c'. wo *gau gau* ta san gungfen
      I tall tall he three centimeter

3.1.2. Reduplicated Action Verbs

Though the categories and argument structures of action verbs after reduplication remain the same, there are more syntactic limitations on reduplicated forms. We observe that these reduplicatives are mutually exclusive with the adjuncts of frequency, quantifier, duration, and postverbal location. Moreover, they are incompatible with the aspects, *le* 'PERFECTIVE', *je* 'DURATIVE', and *guo* 'EXPERIENTIAL' as well as the *bei* construction. As described in section 2.1, reduplication of action verbs denotes the tentative and delimitative aspects which mean some actions are regarded as a piecemeal fashion internally, so any expressions to signal the instantaneous completion of an action, such as the occurrence of aspect, *le* 'PERFECTIVE' and *guo* 'EXPERIENTIAL' certainly violate the semantic function of reduplication. Accordingly, a postverbal locative phrase which refers to the place where an action is achieved is also forbidden. Again, the incompatibility between *bei* construction and reduplication is because *bei* constructions interpret the event in its totality, contrary to the internal event-structure of the tentative aspect[5]. Beside the contradiction of semantic effects, the exclusion of some expressions may be because of the redundancy in meaning. For example, the durative aspect, *je* 'DURATIVE', is suggested by Li and Thompson [6] to express "ongoing, or durative nature of an event' which has been conveyed by reduplicated forms. In addition, tentative reduplicated constructions may also pertain the quantitative meaning of an action which the adjuncts of frequency and quantifier tend to express. It is not necessary to contain two or more expressions which are the same. Therefore, reduplicated forms do not occur with the durative aspect, adjuncts of duration, frequency and quantifier. These restrictions of co-occurrence will be illustrated by the following examples.

---

5. Li and Thompson [6] stipulates that a *bei* sentence "describes an event in which an entity or person is dealt with, handled, or manipulated in some way." This definition entails that the event is considered as a whole.

(12) a. ching *tzuo* <u>yihuei</u>   (duration)
   please sit a while
   'Please sit for a while'

*a'. ching *tzuo tzuo* <u>yihuei</u>
   please sit sit   a while

 b. *tzuo* <u>tzai shrtou shang</u> (postverbal location)
   sit in   stone above
   'Sit on the stone'

*b'. *tzuo tzuo* <u>tzai shrtou shang</u>
   sit sit   in   stone above

 c. *tzou* <u>yi bian</u>   (frequency)
   walk one time
   'Walk once!'

*c'. *tzou tzou* <u>yi bian</u>
   walk walk one time

 d. *tauluen* <u>guo</u> je ge wenti
   discuss EXP this CL question
   'This question has been discussed'

*d'. *tauluen tauluen* <u>guo</u> je ge wenti
   discuss discuss EXP this CL question

3.2. Representation of Reduplication in ICG

Following the above discussion of grammatical feature of reduplication, this section proposes ICG representations of reduplication for efficient parsing. The representation includes syntactic information, such as category, basic patterns (BP) and adjunct precedence (AP), and semantic information, such as semantic features. According to the observation described in section 3.1.1 and 3.1.2, the co-occurrence restriction of adjuncts will be denoted with AP, as in (13).

(13) a. adjunct restriction in Stative verb

syn : constraints: *AP : A1 : v[+rd] > < {degree, comparison}*

b. adjunct restriction in Action Verb

syn: constraints: *AP: A2: v[+rd] > < {complement[ASP],*
frequency, quantifier, duration,
agent[{PP[bei],P[bei]}]}

As for the other information, (a) once recognized as reduplicated constructions, reduplicatives will inherit the same syntactic categories as their original forms, even though they are not stored in the lexicon; (b) reduplicated stative verbs will acquire the feature [+vivid] and reduplicated action verbs, [+tentative]; (c) the feature [+rd] will be specified, while reduplication is identified. Once the [+rd] is specified, A1 in stative verb and A2 in action verb will be applied during the process of parsing language. For further explanation, the stative verb, *kuaile* 'happy' and the action verb, *da* 'hit' and their reduplicated forms are taken as examples.

(14) a. *kuaile*

sem : meaning : happy
   feature : +manner
   adjuncts : ...

syn : class   : VH21
   features :
   constraints: form
      BP B1 : experiencer < * ;

a'. *kuai kuai le le*

sem : meaning : be very happy
   feature : +manner, +*vivid*
   adjuncts : ...

```
syn : class       : VH21
      features   : +rd
      constraints: form
                   BP  B1 : experiencer < * ;

                   AP  A1 : v[+rd] > < {degree, comparison}
```

b. *da*

```
sem : meaning  : hit
      feature  :
      adjuncts : ...

syn : class       : VC2
      features    :
      constraints: form
                   BP  B1: agent[{NP,PP[you]}] < * < goal[NP];
                       B2: agent[{NP,PP[you]}] < goal[PP] <*;
                       B3: goal[NP] < agent[{PP,P[bei]}] < *;
```

b'. *da (yi) da*

```
sem : meaning   : hit a little
      feature   : +tentative
      adjuncts  : ...


syn : class       : VC2
      features    : +rd
      constraints : form
                    BP  B1: agent[{NP,PP[you]}] < * < goal[NP];
                        B2: agent[{NP,PP[you]}] < goal[PP] < *;

                    AP   A2: v[+rd] > < {complement[ASP],
                             frequency, quantifier, duration,
                             agent[{PP[bei],P[bei]}]}
```

From the above examples, both reduplicated expressions are specified by the feature [+rd] in the syntactic feature. And, try to compare (14b) and (14b'): according to the co-occurrence restriction of adjuncts, action verbs with the feature [+rd] are incompatible with *bei* construction, thus example (14b') does not contain the third basic pattern.

IV. Concluding Remarks

In this paper, we present not only the scope and types of reduplicatives but also a set of formation rules to enhance our word segmentation module. Based on the syntactic constraints of reduplicated constructions, we express the co-occurrence restriction of adjuncts in ICG to help parsing. In addition to context-free rules as in the formation of determinative-measure compounds, context-sensitive rules to construct reduplication are required. However, these context-sensitive rules for reduplicatives are now implemented by context-free rule augmented with conditional checks and do not pose any problem for parsing efficiency.

REFERENCES

[1] Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

[2] Chen, Keh-Jiann and Chu-Ren Huang. 1990. "Information-based Case Grammar." *Proceedings of COLING' 90*. pp.54-59.

[3] Chen, Keh-Jiann and Liu Shing-Huan. 1992. "Word Identification for Mandarin Chinese Sentence." *Proceedings of COLING' 92*. Vol.1. pp.101-107. Nantes, France.

[4] Huang, Chu-Ren. 1992. "Adjectival Reduplication in Southern Min: A Study of Morpholexical Rules with Syntactic Effects." In *Chinese Languages and Linguistics. Vol.1. Chinese Dialects*. 407-422. Academia Sinica, Taipei.

[5] Huang, Chu-Ren and Wei-Mei Hong. 1992. "Coordination and Dependency: A Study of Mandarin Comparative *bi*." Paper presented in *the First International Conference on Chinese Linguistics (ICCL I)*, Singapore. June 24-26, 1992.

[6] Li, Charles N. and Sandra A. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.

[7] Mo, Ruo-ping Jean, Yao-Jung Yang, Keh-Jiann Chen and Chu-Ren Huang. 1991. "Determinative-Measure Compounds in Mandarin Chinese." *Proceedings of ROCLING IV*. pp.111-134.