A Computational Study of the *Ba* Resultative Construction: Parsing Mandarin *Ba* Sentences in HPSG

Picus Sizhi Ding

Institute of Linguistics
Academia Sinica, Taipei, Taiwan, R.O.C.
Email: picus@sinica.edu.tw

ABSTRACT

Recognizing Mandarin *ba* sentences as the *Ba* Resultative Construction, this study implements a rigid test of Ding's (1993) treatment of *ba* as a matrix verb in the periphrastic resultative construction. The computational experiment is facilitated with a formal rendition of the analysis of *ba* in Head-driven Phrase Structure Grammar. Moreover, the small parsing experiment also highlights the important role of semantics in natural language processing, including common-sense knowledge.

1. INTRODUCTION

Probably no other lexeme in Mandarin evokes as much controversy as *ba* in regard to the lexical category. The category of *ba* in Modern Mandarin has been claimed to be various types of verbs (including auxiliary verb) as well as a 'deverbalized' preposition (cf. Wang 1947, Chao 1968, Hashimoto 1971, Huang 1974, Li & Thompson 1981, Sun 1996, and Sybesma 1999, among others). What counts towards the so-called '*ba*-construction' is also controversial (for details, see Ding 1993). Mangione (1982) presents an investigation of *ba* sentences most comparable to the one undertaken here in terms of the types of *ba* sentences and the explicit formal approach. The largest divergence lies in his treatment of *ba* as a preposition, heading a sort of 'adverbial prepositional phrase' (Mangione 1982: 158). Since Montague Grammar allows prepositions to head predicates semantically (cf. Dowty et al 1981: 243), the crux of the *ba* problem is dissolved in this framework. Taking advantage of this conflation of verb and preposition, Mangione is able to offer a formal analysis of the predicate headed by *ba* while arguing for treating *ba* as a preposition. This convenience, however, is intrinsically built upon the premises of Montague Grammar.

The perennial debate about the syntactic category of *ba* endures such a lengthy period partly due to neglect of its resultative meaning, and in part because of the discrete-category classification that hinges on a clear-cut distinction. This study aims at demonstrating the advantage of treating *ba* as a resultative verb, proposed in Ding (1993), through a rigid computational implementation of the analysis of the *Ba* Resultative Construction within the framework of Head-driven Phrase Structure Grammar (HPSG). It also includes a small parsing experiment for processing the *Ba* Resultative Construction, showing the crucial interplay between syntax and semantics.

The major parts of the paper are organized as follows: A working definition of *ba* sentences is proffered in §2. Discussing the option of considering *ba* as a resultative verb, §3 analyzes the general syntactic structure of the resultative construction. Following the HPSG analysis presented in §4, §5 is devoted to the parsing experiment that implements the proposed analysis.

2. THE *BA* RESULTATIVE CONSTRUCTION

The following is a working definition for the *Ba* Resultative Construction (BRC, henceforth), advanced in Ding (1993):

> A *ba* sentence belongs to the *Ba* Resultative Construction if, and only if, the object of *ba* holds a proper semantic relationship with the successive clause that denotes a resultative

state. The semantic relation between the object of *ba* and the clausal complement can be PATIENT-and-resultant or EXPERIENCER-and-stative.

Following Nedjalkov & Jaxontov (1988: 6), the notion of 'resultative' is extended to encompass the stative (a state without implication of its origin) in a broader sense. 'Resultative' in its narrow sense is on par with Wang's (1947) 'disposal' inasmuch as both action and state are taken into account. While the term 'resultative' is adapted in its broad sense in this study, 'resultant' will be used to refer to the narrow sense of the resultative, thus 'resultative' = 'resultant' + 'stative'.

According to the definition above, all the sentences below belong to the *Ba* Resultative Construction. Besides the typical *ba* sentence in (1), the resultative construction also covers those with a 'retained' object, as in (2), and those with a causative meaning, e.g. (3).

(1) The regular type of BRC
     da huo   ba   fangzi   gei   shao   le.
     big fire   RV   house   PS   burn   AP
     'The big fire has burned the house (down).'

(2) The object-retained type of BRC
     ta   ba   san   ge   taozi   chi   le   liang   ge.
     s/he   RV   three   CL   peach   eat   AP   two   CL
     'S/He has eaten two out of the three peaches.'

(3) The causative type of BRC
     ni   ba   women   ji   de   budeliao.
     you   RV   we   anxious ET   excessively
     'You have made us excessively distressing.'

On a closer look at the examples (1)-(3), each of the unique characteristics of these sentences is discernible. In the regular type, *ba* shares its two NP arguments with the lexical verb, i.e. the subject of *ba* in (1) is understood as the subject of 'to burn', and the object of *ba* is interpreted to be identical to the object of the lexical verb, as well. While *ba* still shares its subject with the lexical verb in the object-retained type, the object is no longer shared. In other words, each of the verbs in (2) has its own object (the object of *ba* is *san ge taozi* 'three peaches', while the object of *chi* 'to eat' is *liang ge* 'two'). Even though the objects of the verbs are different, some kind of connection holds between them. Semantically, the second object must be associated to the first object in such a way that a resultant state can be set up or construed. The argument sharing is quite different in the causative type. The non-parallel coreference between the object of *ba* and the subject of the lexical verb in (3) is not found in the preceding examples. The dual grammatical role on the NP argument, where the object of *ba* is realized as the subject of the small clause, engenders a causative meaning in this type of *ba* sentence.

The resultative state is crucial to the grammaticality of a *ba* sentence. A well-formed *ba* sentence becomes unacceptable when the resultative meaning is eliminated, as exemplified in (4). Notice that the only difference between (4)a and (4)b is the presence/absence of the perfective aspect marker *le*. Its use in (4)a signifies the completion of an action, and thus entails a resultative state. This resultative meaning vitally affects acceptability of the sentence.

(4)a   shushu   ba   fangzi   mai   le.
        uncle   RV   house   sell   PF
        'Uncle has sold the house.'

   b   *shushu   ba   fangzi   mai.
        uncle   RV   house   sell

3. THE CATEGORY OF *BA* IN MODERN MANDARIN

Recognizing the subtle resultative meaning, Ding (forthcoming) propounds a resultative path hypothesis in addition to the relatively well-establlished serial verb path for the grammaticalization of
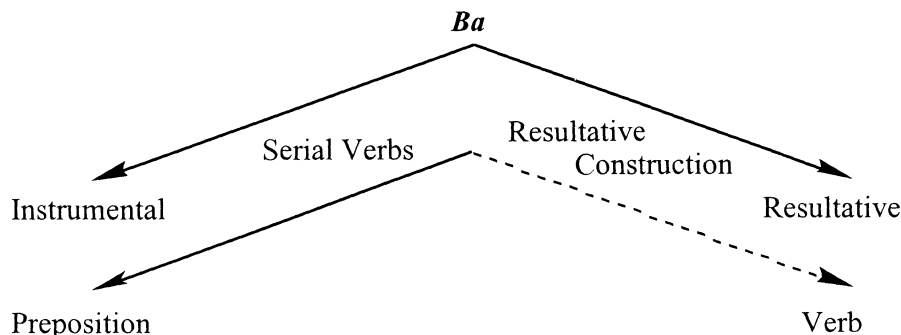
ba, as portrayed in Figure 1.

Figure 1: Bifurcation in the grammaticalization of *ba*

The two-pathed hypothesis accounts for the instrumental use of *ba* in Old Mandarin and the development of the resultative function of *ba* in Modern Mandarin. On the left is the pathway involving serial verbs. It leads to the grammaticalization of *ba* as a 'deverbalized' preposition; the semantic content of the verb has been reinterpreted as instrumental. On the other hand, the path for resultative construction fundamentally represents a semantic change, and not a categorial change. It explains the outcome of the grammaticalization in terms of semantic shift of the domain: from the concrete domain of 'to hold (something)' to the abstract domain of resultative. Whether the change subsequently triggers a categorial shift is determined by other linguistic factors such as further weakening of the semantic content to vacuity. This potential future development is not predicted with the semantic change concerned in this path of grammaticalization. Therefore, *ba* may be treated as a resultative verb in the spectrum from verb to other related categories.

That resultative is a functional meaning does not necessarily ensue that the category of *ba* would logically be functional. Consider the following sketch of verbhood scale in Mandarin:
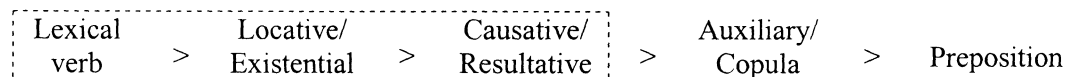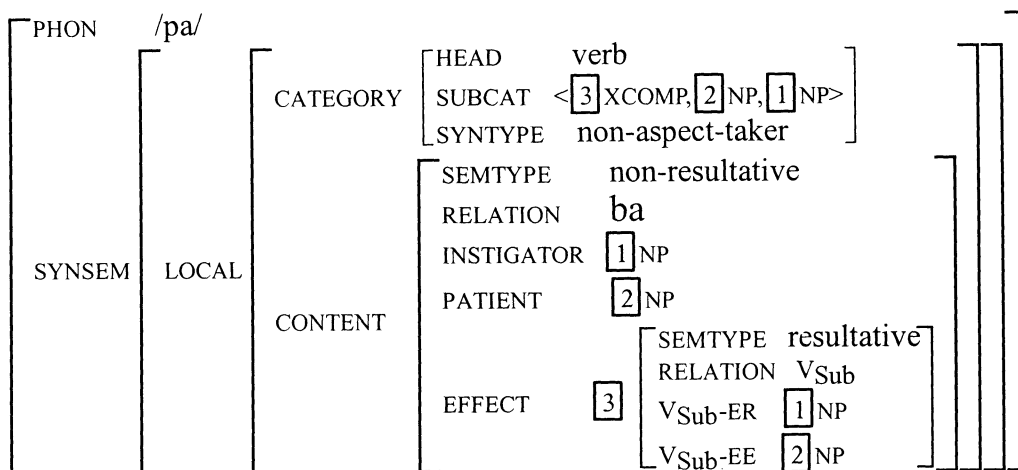

Figure 2: Grammaticalization continuum from verbs to prepositions in Mandarin

The scale is arranged in accordance of semantic contents from concrete to more abstract domains, and thus Lexical verb and Preposition each occupy one end of the scale. These two categories represent the word classes of Verb and Preposition, respectively. Category member-ship of those between the two is rather fuzzy. General diagnostic tests are of little help to them. For instance, the aspect-taking test only identifies lexical verbs in Mandarin. To cope with these problematic categories, a plausible resolution is to extend the membership of the word class of Verb to include locative/existential verbs, and causative/resultative verbs by virtue of their distinguishable (functional) meanings, as indicated by the box. Auxiliary verbs and Copula can form a group either as a subclass under the expanded Verb class or simply as a new word class.

Since resultative is a highly abstract meaning, one could still insist on treating *ba* as a preposition. There is little problem in advocating such a functional category for *ba* in the regular type of BRC. Since there are exactly two NP arguments, the object of *ba* can be regarded as a fronted object of the lexical verb, and the subject simply the agent of the verb (cf. Chao 1968). The inadequacy of this treatment becomes apparent when the other types of BRC are brought into picture. In the object-retained type, the existence of three NP arguments is clearly beyond the subcategorization capacity of the lexical verb, cf. (2) above. Although a counterpart of the *ba* sentence with a postverbal object is sometimes possible for the causative type, this is feasible only when the lexical verb is a resultative verb compound. For instance, the *ba* sentence of the causative type in (5)a cannot be changed to ordinary sentences by rendering the object of *ba* as the object of the lexical verb, as attempted in (5)b.[1]

---

[1] When the original object of *ba* (*mama* 'mother') serves as the subject of *lei* 'to be tired' — not the

With these difficulties, the treatment of *ba* as a preposition is unsatisfactory.

(5) a haizide   yifu   ba   **mama**   *xi*   de   lei   ji   le.
        child's   clothes   RV   mother   wash   ET   tired extremely PF
        'Washing the children's clothes has got Mother extremely tired.'

   b *haizide   yifu   *xi*   **mama**   de   lei   ji   le.
        child's   clothes   wash   mother   ET   tired extremely PF
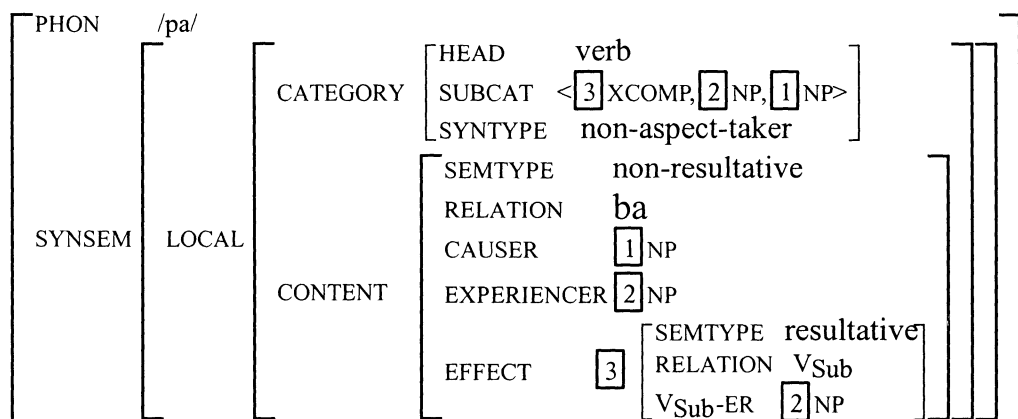
Even though grammaticalized, *ba* has retained some verbal properties in BRC (cf. Ding 1993; forthcoming, for details). Thus it is feasible to consider the resultative verb syntactically as a verb. This alternative approach then leads to a unified subcategorization for the resultative verb, regardless of the type of BRC:

(6)   Resultative *ba* = Tri-argument verb: Two NPs
                                                  One subordinate clause with resultative content

The relation of the resultative verb and its arguments is depicted in Figure 3. *Ba* essentially assigns three semantic roles to its arguments: INSTIGATOR, PATIENT and EFFECT. This is readily observable from the regular and the object-retained types of BRC. The semantic relation between the object of *ba* and the verbal complement in these cases is PATIENT-and-resultant. In the case of causative type, the first two semantic roles are assigned respectively as CAUSER and EXPERIENCER, with the semantic relation of EXPERIENCER-and-stative between the object of *ba* and the complement clause.

$$\text{Subject} \;\leftarrow\!\!-\!\!- \begin{bmatrix} \text{INSTIGATOR} \\ \text{CAUSER} \end{bmatrix} -\!\!-\!\!- \textbf{ba} -\!\!-\!\!- \begin{bmatrix} \text{PATIENT} \\ \text{EXPERIENCER} \end{bmatrix} -\!\!-\!\!\rightarrow \text{Object}$$

[EFFECT]

↓

Verbal Complement

Figure 3: The grammatical relation between *ba* and its arguments

4. AN ANALYSIS OF THE *BA* RESULTATIVE CONSTRUCTION IN HPSG

The essence of the syntactic analysis of the *Ba* Resultative Construction to be presented below remains unchanged from that in Ding (1993). Adopting Government and Binding Theory as the framework, Ding (1993) is unable to avoid the notions of case, tense, and trace, all in the null form, which renders a grammar of Mandarin astonishingly similar to that of English. The change of framework to HPSG makes it possible to discuss the grammar of Mandarin outside the shadow of English. The abstract case and tense can be spared, and the notion of trace is coped with in terms of structure sharing between NPs.

As a resultative verb, *ba* always takes two NPs and a verbal complement. The differences between the three types of *ba* sentence are materialized in the peculiar way in which the NP arguments are shared between *ba* and the lexical verb. The Attribute Value Matrix (AVM) of *ba* in each type of BRC is provided in (7). Notice that the specification of CATEGORY in each AVM is exactly identical. The divergence between these AVMs is primarily found within the part EFFECT, in which indexes of NPs differ from one another.

---

object of *xi* 'to wash', it is possible to paraphrase (5)a as follows:

   haizide   yifu   *xi*   de   **mama** lei   ji   le.
   child's   clothes   wash   ET   mother   tired extremely PF
   'Washing the children's clothes has got Mother extremely tired.'

(7)a The AVM of *ba* in the regular type of BRC[2]

$$
\begin{bmatrix}
\text{PHON} & \text{/pa/} \\
\text{SYNSEM} & \begin{bmatrix}
\text{LOCAL} & \begin{bmatrix}
\text{CATEGORY} & \begin{bmatrix}
\text{HEAD} & \text{verb} \\
\text{SUBCAT} & \langle \boxed{3}\text{XCOMP}, \boxed{2}\text{NP}, \boxed{1}\text{NP} \rangle \\
\text{SYNTYPE} & \text{non-aspect-taker}
\end{bmatrix} \\
\text{CONTENT} & \begin{bmatrix}
\text{SEMTYPE} & \text{non-resultative} \\
\text{RELATION} & \text{ba} \\
\text{INSTIGATOR} & \boxed{1}\text{NP} \\
\text{PATIENT} & \boxed{2}\text{NP} \\
\text{EFFECT} & \boxed{3}\begin{bmatrix}
\text{SEMTYPE} & \text{resultative} \\
\text{RELATION} & \text{V}_{\text{Sub}} \\
\text{V}_{\text{Sub}}\text{-ER} & \boxed{1}\text{NP} \\
\text{V}_{\text{Sub}}\text{-EE} & \boxed{2}\text{NP}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

b The AVM of *ba* in the object-retained type of BRC

$$
\begin{bmatrix}
\text{PHON} & \text{/pa/} \\
\text{SYNSEM} & \begin{bmatrix}
\text{LOCAL} & \begin{bmatrix}
\text{CATEGORY} & \begin{bmatrix}
\text{HEAD} & \text{verb} \\
\text{SUBCAT} & \langle \boxed{3}\text{XCOMP}, \boxed{2}\text{NP}, \boxed{1}\text{NP} \rangle \\
\text{SYNTYPE} & \text{non-aspect-taker}
\end{bmatrix} \\
\text{CONTENT} & \begin{bmatrix}
\text{SEMTYPE} & \text{non-resultative} \\
\text{RELATION} & \text{ba} \\
\text{INSTIGATOR} & \boxed{1}\text{NP} \\
\text{PATIENT} & \boxed{2}\text{NP}* \\
\text{EFFECT} & \boxed{3}\begin{bmatrix}
\text{SEMTYPE} & \text{resultative} \\
\text{RELATION} & \text{V}_{\text{Sub}} \\
\text{V}_{\text{Sub}}\text{-ER} & \boxed{1}\text{NP} \\
\text{V}_{\text{Sub}}\text{-EE} & \boxed{4}\text{NP}*
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

c The AVM of *ba* in the causative type of BRC

$$
\begin{bmatrix}
\text{PHON} & \text{/pa/} \\
\text{SYNSEM} & \begin{bmatrix}
\text{LOCAL} & \begin{bmatrix}
\text{CATEGORY} & \begin{bmatrix}
\text{HEAD} & \text{verb} \\
\text{SUBCAT} & \langle \boxed{3}\text{XCOMP}, \boxed{2}\text{NP}, \boxed{1}\text{NP} \rangle \\
\text{SYNTYPE} & \text{non-aspect-taker}
\end{bmatrix} \\
\text{CONTENT} & \begin{bmatrix}
\text{SEMTYPE} & \text{non-resultative} \\
\text{RELATION} & \text{ba} \\
\text{CAUSER} & \boxed{1}\text{NP} \\
\text{EXPERIENCER} & \boxed{2}\text{NP} \\
\text{EFFECT} & \boxed{3}\begin{bmatrix}
\text{SEMTYPE} & \text{resultative} \\
\text{RELATION} & \text{V}_{\text{Sub}} \\
\text{V}_{\text{Sub}}\text{-ER} & \boxed{2}\text{NP}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

---

[2] Shorthands used in the matrix diagram include: PHON (phonetic form), SYNSEM (syntactic and semantic features), SUBCAT (subcategorization list), XCOMP (verbal complement), SYNTYPE (syntactic type), SEMTYPE (semantic type), V_{Sub} (subordinate verb), $\text{V}_{\text{Sub}}\text{-ER}$ (subject of subordinate verb), and $\text{V}_{\text{Sub}}\text{-EE}$ (object of subordinate verb).

In (7)a, the structure sharing is straightforward. All the NP arguments are shared in parallel between *ba* and its subordinate verb; the subject of *ba* is identical to the subject of the subordinate verb, and the object of *ba* is also the object of the subordinate verb. On the other hand, (7)b shows a structure-sharing between the subjects of *ba* and its subordinate verb only; the objects are not shared. Nonetheless, these two objects must hold a kind of semantic relation, indicated by the sign ⊛ . This semantic relation can be realized in an array of relationships such as whole-and-part, possessor-and-possessee, and so forth.[3] Finally, with the coreference between the object of *ba* and the subject of the subordinate verb, (7)c leads to a causative interpretation. Also note that the embedded verb in this type takes only an NP for subject argument, without any object argument.

Two non-standard HPSG specifications in the matrix diagrams are: SYNTYPE and SEMTYPE. Being an aspect language, Mandarin has a constraint as to whether a verb can take an aspect marker. Within the enlarged verb class adopted for Mandarin (cf. Figure 2), a small number of verbs cannot take any aspect markers. *Ba*, used as a resultative verb, belongs to this group, along with the casuative verb *rang*, the locative verb *zai*, and the existential *you*. To establish this restriction, the specification of SYNTYPE is created under CATEGORY of verbs. The SYNTYPE functions like a parameter with a binary value of 'aspect-taker' and 'non-aspect-taker'.

The devising of SEMTYPE is mainly inspired by the semantic classification of verbs in Chafe (1970).[4] The new specification is attached to verbs and nouns as part of their CONTENT. Below we will discuss the SEMTYPE of verbs. Section 5.1 will look at the SEMTYPE of nouns.

It is well-observed that verbs can be divided into two major groups in accordance with their meanings, i.e. action and state. Chafe (1970: 95-104) proposes a schema with the features of 'state', 'process', and 'action' for systematic classification of English verbs.[5] Some examples of English verbs and their classification are given in (8):

(8) a   State   The wood is **dry**.
    b   Process   The wood **dried**.
    c   Action   Mother **laughed**.
    d   Process-Action   Mother **dried** the wood.

It must be noted that Chafe's classification is not a purely semantic one. To some extent, it unavoidably involves syntactic features of a lexeme. This is certainly true in the case of 'state' predicate in English, which essentially contains an adjective. It is also feasible to predict a specific kind of predicate, simply based on syntactic criteria. For instance, a predicate consisting of an adjective in English entails that the predicate must have the feature of 'state'. The major groups of Mandarin verbs, following Chafe's schema of classification, are exemplified as follows:

(9) a   State   mukuai shi **gan** de. 'The wood is dry.'
    b   Process   mukuai **gan** le. 'The wood dried.'
    c   Action   mama **xiao** le. 'Mother laughed.'
    d   Action-Process   mama **hong-gan** le mukuai. 'Mother dried the wood by heating it.'

Chafe (1970: 132) also introduces a variety of derivational rules such as causative, inchoative, and resultative, etc. These rules allow derivational processes to take place between verbs of sundry classes. Taken as a semantic notion, the term 'resultative' in this paper, however, is adopted in a sense fundamentally different from that used by Chafe.

Given its semantic nature, the concept of 'resultative' is not confined to verb at the lexical level; it

---

[3] If the two object NPs in question are perceived as each composing a set of semantic features, the two sets must have some salient elements in common so as to form an acceptable *ba* sentence. The criterion for being a salient element is determined by whether or not the element can function as a bridge to allow a resultative meaning 'propagated' to the matrix object from the embedded object upon which the effect originally acts.

[4] This idea is further reinforced with Paul McFetridge's adoption of Roget's thesaurus (cf. Chapman 1992) in an on-going machine translation project.

[5] Chafe (1970) extends the notion of 'verb' to include predicative adjectives.

is equally applicable to a verb phrase. Although *ba* is a resultative verb, the resultative meaning does not reside in the lexical content. Thus in the matrix diagrams of *ba*, the resultative verb is lexically specified as 'non-resultative' in SEMTYPE. On the other hand, the resultative verb requires that the verbal complement as a whole possess 'resultative' in SEMTYPE regardless of the lexical specification of SEMTYPE for the embedded verb. In other words, SEMTYPE may be used as a device to signify whether a resultative meaning is present in a verb phrase.

The specification of SEMTYPE has not been elaborated to a multi-level hierarchy. In the simplest form, the inherited hierarchy consists of two levels. For categories other than noun/pronoun, at the top level are 'resultative' versus 'non-resultative', as outlined in the following:[6]

Class	Non-resultative			Resultative	
Division	State	Process	Action	Action-process	Perfective
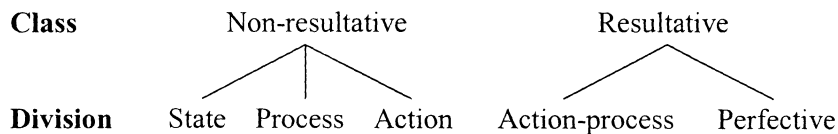
Figure 4: A semantic classification of Mandarin verbs

At the bottom level, various divisions are subsumed exclusively under either of the classes, 'resultative' or 'non-resultative'. The 'non-resultative' class includes verbs of 'state', 'process' and 'action'. Verb compounds of 'action-process', however, have the SEMTYPE 'resultative'. Another subgroup under 'resultative' is 'perfective'. Since the perfective aspect implies a change of state/situation, which can easily be construed as featuring 'resultative'.

The HPSG analysis of BRC submitted here is capable of coping with the interface between syntax and semantics in a rather straightforward manner. With the unified subcategorization of *ba*, the implementation of the analysis is much facilitated. The remainder of the paper will concentrate on parsing the resultative construction, based on the proposed analysis.

5. PARSING THE *BA* RESULTATIVE CONSTRUCTION

In an ideal natural language processing model, all aspects of linguistic data must be taken into consideration. Linguistic representations at varied levels bear important factors in language processing. HPSG in its current stage has embraced different aspects of language, notably syntax, semantics, and pragmatics (cf. Pollard & Sag 1994). In addition, as a declarative and constraint-based grammar, HPSG renders itself a good candidate for providing a theoretical framework in developing an ideal parser. In light of these benefits, the parsing experiment below has selected to carry out within the framework of HPSG.

5.1 The Parser

Since the goal of the parsing experiment is to test the rigidity of the proposed analysis of *ba* rather than examining the performance of a parser, only essential details of the parser will be given here. With the foregoing HPSG analysis of the resultative verb *ba*, the AVMs can easily be translated and integrated to a parser. The parser used in the test is not deliberately designed for this experiment; it is adapted from a parsing program written for parsing simple English sentences in HPSG.[7] The original program consists of four parts: *Grammar*, *Knowledge*, *Lexical Hierarchy*, and *Lexicon*. Keeping the basic algorithm intact, the adapted program for processing Mandarin *ba* sentences has retained the number and structure of the subprograms. The revised subprograms comprise *Grammar*, *Mandarin Lexical Hierarchy*, *Cooking Knowledge*, and *Cooking Lexicon*.

The subprogram *Grammar* has remained intact for the Mandarin parser, since it deals with principles and rules pertaining to HPSG. Given that the framework is identical, there is no need to

---

[6] The divisions shown here are only those relevant to the parsing experiment. They are by no means exhaustive.

[7] This is only an elementary parser, capable of parsing short and simple English sentences. According to the original designer, Paul McFetridge (personal communication), the parser is based on a commonly used algorithm such as the one discussed in Earley (1970).

modify this component of the program. On the contrary, it is necessary to replace contents of the other subprograms so that it parses Mandarin instead of English.

The *Lexical Hierarchy* is responsible for the morpho-syntax of a particular language. Grammatical relations in Mandarin are provided in a new version of this component. Compared with the original subprogram, the *Mandarin Lexical Hierarchy* has undergone a considerable degree of 'simplification' because of the lack of morphological inflection in Mandarin. Whereas a noun in English needs to specify its case to conform with the overt case-marking of pronouns; nouns in Mandarin can dispense with this specification. A similar treatment is also seen on plurality of nouns: no specification is attempted for plurality in the *Mandarin Lexical Hierarchy*.[8] This strategy of underspecification is also applied in respects to agreement and tense. Since Mandarin is an aspect language, a brief section on aspect marker is included. A major addition results from the incorporation of the semantic classification of Mandarin verbs shown earlier in Figure 4.

Another noticeable modification concerns the determiner. Since the use of articles with nouns is common in English, NPs in the original *Lexical Hierarchy* are established in such a manner that the head noun of the NP subcategorizes for an article as its determiner. This elaborated treatment of NPs does not seem to be necessary in Mandarin. In spite of the frequent use of classifiers with nouns in Mandarin, classifiers do not play a role as essential as that of articles to nouns in English. The use of classifiers presupposes that the reference of the target noun is construed in a relatively concrete sense, in which the noun can be counted or measured. In light of this, the *Mandarin Lexical Hierarchy* simply marks nouns as lexical. When a classifier qualifies a noun, it can be handled as an adjunct such as an adjective phrase.

The cooking domain is chosen for the parsing experiment because food preparation essential-ly entails change of state of ingredients, especially in Chinese cooking. The nature of cooking thus provides appropriate contexts for the occurrence of *ba* sentences. Moreover, the size of the domain in terms of lexicon can be flexibly increased or reduced, depending on the complex level of an experiment. The cooking domain involved here is very small, but functionable.

A minimal knowledge about cooking is supplied in the *Cooking Knowledge*. The information deals with what is generally called 'common sense'. To the human mind, it is common sense that one does not cut salt, but often needs to cut vegetable and meat into smaller pieces. This rather trivial piece of information, however, needs to be stipulated clearly in order to equip the parser with some artificial intelligence. Within the domain of cooking, efforts are made for the parser to determine 'intelligently' whether a given sentence is legitimate, i.e. parsable (precluding such contexts as fable stories and science fictions). To this goal, semantic classification of nouns is called for. As with verbs, the SEMTYPE of nouns consists of two inherited hierarchical levels. The overall structure of the classification for nouns involved in this experiment is given as follows:
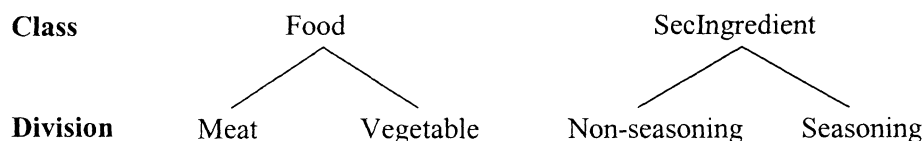
Class	Food		SecIngredient	
Division	Meat	Vegetable	Non-seasoning	Seasoning

Figure 5: A semantic classification of cooking-related nouns

The class of 'Food' consists of 'Meat' and 'Vegetable', whereas the class of 'SecIngredient' (secondary ingredient) 'Non-seasoning' and 'Seasoning'. In addition, it is expressly required that the object of 'to cut' have the SEMTYPE value of 'Food'. With all this information properly encoded in the *Cooking Knowledge*, the parser is capable of distinguishing abnormal utterances like 'cut the oil' from logical sentences in the cooking domain.

Finally, the *Cooking Lexicon* comprises lexical entries with 'terminology' defined primarily in the *Mandarin Lexical Hierarchy*. Since each lexical item must be declared in the *Cooking Lexicon*, it offers an overview of lexemes used in the experiment. For instance, the lexical entries of *ba* for the regular and the object-retained types are stated as in (10)a and (10)b, respectively:

---

[8] Although Mandarin has a pseudo-plural marker *men*, its use with animate nouns heavily depends on pragmatics. This marker is not dealt with here.

(10)a

> (lex ba ((phon) = ba
> ResultativeVerb
> (synsem local category intsubcat rest first synsem local category lex) = +
> (synsem local category intsubcat rest first synsem local category subcat first)
> = (synsem local category subcat first)
> (synsem local category intsubcat rest first synsem local category subcat rest
> first) = (synsem local category subcat rest rest first)))

b

> (lex ba ((phon) = ba
> ResultativeVerb
> (synsem local category intsubcat rest first synsem local category lex) = -
> (synsem local category intsubcat rest first synsem local category extsubcat)
> = (synsem local category subcat rest rest first)))

Note that the entry in (10)b is shorter than that in (10)a. This is because no declaration is made in (10)b for the coreference between the matrix and the embedded objects in the object-retained type of BRC.[9] The term 'ResultativeVerb' is defined in the *Mandarin Lexical Hierarchy*, based on the unified subcategorization of resultative *ba* discussed earlier; it runs as follows:

(11)

> (inherit ResultativeVerb
> (Non-resultative Trueverb Object2 Subject-third
> (synsem local content instigator) = (synsem local category extsubcat synsem local content index)
> (synsem local content patient) = (synsem local category intsubcat first synsem local content index)
> (synsem local content effect) = (synsem local category intsubcat rest first synsem local content)
> Non-aspect-taker
> (synsem local category intsubcat rest first synsem local category head sort maj) = verb
> (synsem local category intsubcat rest first synsem local content semtype class) = resultative
> (synsem local category intsubcat rest first synsem local category head sort aux) = -))

Other entries in the *Cooking Lexicon* include 11 nouns (such as *yu* 'fish', *you* 'oil', *yan* 'salt', etc.), 8 concrete verbs (*jia* 'add', *zha* 'deep-fry', *chao* 'stir-fry', *qie* 'cut', *re* 'heat', *qu* 'drain', *banyun* 'mix well', and *nonghu* 'thicken'), 1 adjective (*zhahaode* 'deep-fried'), and the perfective aspect marker *le*. The *Cooking Lexicon* is undoubtedly very small in size, but it is sufficient for us to produce many grammatical as well as problematic *ba* sentences for testing.

5.2 The Parsing Results

A session of running the parser is presented here to show what the experimental program can and cannot process. While the results of the parsing help to illustrate how the principles in theory can be implemented, they also exhibit limitations of the parser. For the ease of discussion, the script of the parsing results is divided and presented in separate pieces. (Performance information such as parsing time is omitted out of space consideration.) When a sentence is unacceptable, the parsing results will indicate that 'Number of parses' is zero, as shown in (12), where *ba* is modified by an aspect marker. The parser is able to detect these abnormalities because it has been expressly stated in the SYNTYPE of *ba* that the verb is 'non-aspect-taker'.

(12)　　*ba　　le　　jia　　le　　tang
　　　　　 RV　　AP　　add　　AP　　sugar

> USER(2): (hpsg "ba le jia le tang")
>
> Number of parses: 0

---

[9] No entry for the causative use of *ba* is included, since this type of BRC hardly occurs in the cooking domain.

Exemplified in (13) are two parsable *ba* sentences. With the presence of the aspect marker *le*, (13)b is slightly more complicated than (13)a. The structure of sentence is displayed in square brackets in response to the command '(show-phon)'. The major semantic contents of the sentences follow after the command '(show-semantics)'. Beginning with the matrix level, the first line displays that the class of SEMTYPE of *ba* is 'non-resultative'; the subject of *ba* is INSTIGATOR, which is unspecified since it is not given; the object of *ba* is PATIENT; and EFFECT contains information about the embedded verb.

(13)a   ba   dianfen   nonghu   b   ba   you   re   le
        RV   starch   thicken   RV   oil   heat AP
        'thicken the starch'   'heated the oil'

```
USER(11): (hpsg "ba dianfen nonghu")

Number of parses: 1

USER(12): (show-phon)

[BA DIANFEN NONGHU]
T
USER(13): (show-semantics)

((SEMTYPE
((CLASS NON-RESULTATIVE)))
(RELATION BA) (INSTIGATOR _2452)
 (PATIENT ((REF #:R290) (PER 3)))
 (EFFECT
  ((SEMTYPE ((CLASS RESULTATIVE)
(DIVISION ACTION-PROCESS)))
   (RELATION NONGHU) (THICKENER
_2452)
   (THICKENED ((REF #:R291) (PER 3))))))
T
USER(14): (show-syntax)

((HEAD ((SORT ((MAJ VERB) (AUX -)))
(INV -)))
 (SUBCAT
  ((REST NONE)
   (FIRST ((SYNSEM
           ((ORDER <)
            (LOCAL
             ((CATEGORY ((HEAD
((SORT ((MAJ NOUN)))))))
              (CONTENT ((INDEX
_2452)))))))))))
 (LEX -))
T
```

```
USER(15): (hpsg "ba you re le")

Number of parses: 1

USER(16): (show-phon)

[BA YOU [RE LE]]
T
USER(17): (show-semantics)

((SEMTYPE
((CLASS NON-RESULTATIVE)))
(RELATION BA) (INSTIGATOR _3161)
 (PATIENT ((REF #:R292) (PER 3)))
 (EFFECT
  ((SEMTYPE ((CLASS RESULTATIVE)))
(FIRST ((RELATION ASPECT)))
   (REST ((RELATION RE) (HEATER _3161)
(HEATED ((REF #:R293) (PER 3)))
        (SEMTYPE ((CLASS NON-
RESULTATIVE) (DIVISION ACTION)))))))))
T
USER(18): (show-syntax)

((HEAD ((SORT ((MAJ VERB) (AUX -)))
(INV -)))
 (SUBCAT
  ((REST NONE)
   (FIRST ((SYNSEM
           ((ORDER <)
            (LOCAL
             ((CATEGORY ((HEAD
((SORT ((MAJ NOUN)))))))
              (CONTENT ((INDEX
_3161)))))))))))
 (LEX -))
T
```

In (13)a, the embedded verb *nonghu* 'to thicken' belongs to the action-process type, and hence its class of SEMTYPE is 'resultative'. In contrast, the embedded verb *re* 'to heat' in (13)b belongs to the process type. Like other simplex verbs, its class of SEMTYPE is 'non-resultative'. For (13)b to be acceptable, the use of the aspect marker *le* is crucial. Note that the verb phrase consisting of the verb *re* 'to heat' and the perfective aspect is marked as 'resultative' for the class of SEMTYPE of the phrase in the line immediately below EFFECT. If the perfective aspect is not available, a *ba* sentence with a simplex verb alone is unparsable. This is shown in (14) with an action verb. If the resultative meaning of *ba* sentences were to be neglected, the unacceptability of (14) would have to be explained by indirect constraints such as non-monosyllable condition for the embedded clause.

(14)	*ba	yan	jia
	RV	salt	add

```
USER(19): (hpsg "ba yan jia")

Number of parses: 0
```

When employed in cooking books, *ba* sentences are typically subjectless. Near the bottom of (13), the block of information introduced by '(show-syntax)' presents the status of the predicate. Notice that *ba* is looking for an NP on the left (indicating by 'ORDER <') to serve as its subject. Note also that the index number of INSTIGATOR and that of subject of the embedded verb are identical to the missing subject, indexed as '2452' in (13)a and as '3161' in (13)b, respectively.

Although the *ba* sentence in (15)a has satisfied the resultative condition with the perfective marker *le*, the sentence is unacceptable due to the meaning '*cut the water'. Thanks to the *Cooking Knowledge*, the parser is able to detect the abnormality of the sentence. Recall that the verb 'to cut' only takes nouns of the class 'food' as its object in the cooking domain. Therefore, neither (15)a nor (15)b is parsable with 'water' being the object of 'to cut'. These examples have shown that the sharing of the object between *ba* and the subordinate verb has been implemented properly in the parser: *Shui* 'water', the object of *ba*, is understood as the object of 'to cut' in (15)a by the parser, and thus the sentence is rejected for the same semantic consideration as in (15)b.

(15)a	*ba	shui	qie	le	b	*qie	le	shui
	RV	water	cut	AP		cut	AP	water

```
USER(20): (hpsg "ba shui qie le")

Number of parses: 0
```

```
USER(21): (hpsg "qie le shui")

Number of parses: 0
```

It should be borne in mind that the information supplied in the *Cooking Knowledge* is deficient. It covers only a very small portion of the cooking domain. Furthermore, it does not consider such pragmatic factors as implicature and presupposition. Consequently, the parser may reject some well-formed *ba* sentences because of its 'ignorance'. For example, (16) is an acceptable utterance with an implicit resultative meaning.

(16)	ba	zhahaode	zhurou	qu	you
	RV	deep-fried	pork	drain	oil

'drain oil out of the deep-fried pork'

From the meaning of the sentence, we can infer that the pork having been deep-fried is saturated with oil. The process of draining will bring about a change of state, and thus the *ba* sentence is fine. Nonetheless, this kind of implicature is beyond the limited intelligence of the parser. As a result, the sentence is rejected. To make it parsable, an explicit resultative meaning is required. With the aspect marker *le* inserted as in (17), the parser accepts the sentence as a grammatical instance of BRC.

(17)	ba	zhahaode	zhurou	qu	le	you
	RV	deep-fried	pork	drain	AP	oil

'drained oil out of the deep-fried pork'

It should also be mentioned in passing that the small parser is not error-free. When parsing a *ba* sentence with a complex NP object such as the one in (17), the parser fails to properly coindex the subject of *ba* with the subject of the subordinate verb (both implicit in the sentence). The problem is of technical nature, probably lying in the source codes of the original parser.

6. CONCLUSION

This paper has adopted an alternative approach — treating *ba* as a resultative verb — to the *Ba* Resultative Construction. The advantage and adequacy of the analysis are demonstrated with the implementation of the HPSG analysis of the construction in a small parser. This treatment has provided a unified analysis of BRC, unavailable from regarding *ba* as an auxiliary verb or as a

preposition (unless within a theory of grammar that lumps verbs and prepositions together). Moreover, recognizing *ba* as a verb, we do not need to deal with the superficial word order change, as claimed in Li & Thompson (1974), which would only unduly complicate the grammar of Modern Mandarin.

The small parser experiment also shows the importance of semantics in processing natural languages. To equip the parser with a minimum artificial intelligence, a deliberate component for basic knowledge about the real world must be included. By the same token, the specification of semantic features cannot be downplayed in declaring the grammatical properties of the resultative verb, given the crucial constructional meaning of BRC.

ABBREVIATIONS IN GLOSSES

| AP | Aspect marker | CL | Classifier | ET | Extent marker |
| PF | Perfective aspect | PS | Passive marker | RV | Resultative Verb |

ACKNOWLEGEMENTS

REFERENCES

[1]   Chafe, W. 1970. *Meaning and the Structure of Language*. Chicago: University of Chicago Press.

[2]   Chao, Y. R. 1968. *A Grammar of Spoken Chinese*. Berkeley, CA: University of California Press.

[3]   Chapman, R. (ed). 1992. *Roget's International Thesaurus*. New York: HarperCollins Publishers.

[4]   Ding, S. 1993. *The Ba Resultative Construction: A Comprehensive Study of Mandarin Ba Sentences*. M.A. thesis, Simon Fraser University.

[5]   Ding, S. (forthcoming). Semantic Change vs. Categorial Change: A study of the development of *ba* in Mandarin. To appear in *Journal of Chinese Linguistics*, 28.

[6]   Dowty D., R. Wall, & S. Peters. 1981. *Introduction to Montague Semantics*. Dordrecht: Reidel.

[7]   Earley, J. 1970. An efficient context-free parsing algorithm. *Communications of the ACM*, 13,2: 94-102.

[8]   Hashimoto, A. Yue. 1971. Mandarin syntactic structures. *Unicorn*, No. 8. Princeton: Princeton University.

[9]   Huang, S. 1974. Mandarin Causatives. *Journal of Chinese Linguistics*, 2,3: 354-369.

[10] Li, C. & S. Thompson. 1974. An explanation of word order change: SVO → SOV. *Foundations of Language*, 56, 2: 251-299.

[11] Li, C. & S. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, CA: University of California Press.

[12] Mangione, L. 1982. *The Syntax, Semantics and Pragmatics of Causative, Passive and Ba Constructions in Mandarin*. Ph.D. dissertation, Cornell University.

[13] Nedjalkov, P. & S. Jaxontov. 1988. The Typology of Resultative Construction. In P. Nedjalkov (ed), *Typology of Resultative Constructions*: 3-62. Amsterdam: John Benjamins.

[14] Pollard, C. and Sag, I. 1994. *Head Driven Phrase Structure Grammar*. University of Chicago Press.

[15] Sun, C. 1996. *Word Order Change and Grammaticalization*. Stanford: Stanford University Press.

[16] Sybesma, R. 1999. *The Mandarin VP*. Dordrecht: Kluwer Academic Publishers.

[17] Wang, L. 1947. *Zhongguo Xiandai Yufa* (A modern grammar of Chinese). Shanghai: Commercial Press.