ChinGram: A TRALE Implementation of an HPSG Fragment of Mandarin Chinese

Stefan Müller
Freie Universität Berlin
Stefan.Mueller@fu-berlin.de

Janna Lipenkova
Freie Universität Berlin
Janna.Lipenkova@fu-berlin.de

Abstract

In this paper, we present our effort in the development of a HPSG grammar for Chinese. We present the basic notions of the HPSG framework, review existing theoretical analyses and implementations of Chinese grammar fragments in HPSG and present a range of deep linguistic analyses that are part of our own implementation.

1 Introduction

This paper presents a grammar fragment of Chinese which is built in the framework of HPSG (Pollard and Sag 1994) and implemented in the grammar development system Trale (Meurers et al. 2002; Penn 2004). The grammar is one of the grammars that are developed in the CoreGram project (Müller 2013a). Apart from the Chinese grammar, which will be documented in Müller and Lipenkova (In Preparation), there are smaller fragments of several languages and larger fragments of German, Persian, Danish, and Maltese (see Müller (2013b) for details on size). These grammars share a common core und hence crosslinguistic generalizations are captured. We see the advantages of the HPSG framework for a formal analysis of Chinese as follows:

- HPSG provides a range of powerful formal tools for the description of linguistic expressions which are embedded into the logical framework of Typed Feature Structure Logic (Carpenter 1992) and allow a seamless implementation in logical programming paradigms.
- HPSG makes restricted use of *a-priori* theory-internal statements about the empiricial properties of linguistic signs. Since Chinese phenomena often cannot be explained using the terminology and assumptions of the Western linguistic tradition, HPSG provides us with a 'neutral' framework for the formalization of language-specific phenomena based on which more general principles can be derived.
- In contrast to most formal theories, HPSG is not a syntax-driven framework. That is, there is no central syntactic component from which a Phonological Form and a Logical Form is derived. Instead, the different levels of linguistic representation – phonology, syntax, semantics, pragmatics – have equal weight. This is especially beneficial for Chinese, which has a poor morphological system and exhibits a high degree of surface ambiguity. The use of a powerful semantic-pragmatic component with fine-grained definitions of semantic types and selectional restrictions and preferences thus helps disambiguation.

In the following, we first introduce the basic feature architecture and formal tools of the grammar formalism. Then, we review existing work in HPSG and grammar development for Chinese. Finally, we describe the theoretical and empirical basis of our research and provide a synopsis of the covered phenomena; the main analytical choices

* The work reported in this paper was supported by the grant ChinGram MU 2822/5-1 by the Deutsche Forschungsgemeinschaft.

** The following abbreviations are used:

- HPSG sign features: HD: head; SS: synsem; IND: index
- Tree arc symbols: Arg: argument daughter; Spr: specifier daughter; H: head daughter; NH: non-head daughter; Adj: adjunct daughter
- Glosses: CL: classifier; ATTR: attributive particle *de*; LOC: localizer particle

27th Pacific Asia Conference on Language, Information, and Computation pages 240—249

are illustrated using a subset of example phenomena.

2 Framework and implementation

This part provides a brief overview of the main principles and components of HPSG; for more detailed expositions, the reader is referred to the standard makeup described in Pollard and Sag (1994), Sag (1997), and Müller (2008b). The semantics follows Minimal Recursion Semantics as described in Copestake et al. (2005).

2.1 HPSG

The main characteristics of the HPSG framework are as follows:

- *Feature-based*: the universal format of representation are descriptions of typed feature structures (Carpenter 1992).
- *Model-theoretic*: generalizations on linguistic objects are formulated as declarative constraints; there are no transformations.
- *Lexicalist*: a great part of linguistic information, especially information about syntactic combination, is stored in the lexicon.
- *Monostratal*: multiple levels of linguistic representation (phonology, syntax and morphology, semantics, pragmatics and information structure) are modelled in parallel; no formal priority is given to the structural levels.

Formally, an HPSG grammar consists of two parts: a signature and a theory, the theory is subdevided into a lexicon and a grammar:

- *Signature*: ontology that contains types and their feature specifications; the signature is structured as an inheritance hierarchy allowing for multiple inheritance.
- *Lexicon* (constraints on linguistic signs of type *root*, *stem*, or *word*):
  - Lexical entries
  - Lexical rules, specifying systematic relationships holding between classes of lexical items (cf. Meurers 1999a and Meurers 2001, *i.a.*, for studies of the formal properties of lexical rules)
- *Grammar* (constraints on linguistic objects of type *phrase*):
  - Small set of broad-range principles holding of large subtypes of *phrase*, e. g. Head Feature Principle, Valence Principle, Semantics Principle for *headed-phrases*
  - Immediate dominance schemata, specifying the constituency of phrases
  - Linear precedence rules, ruling out impossible constituent orders

A linguistic sign is modelled with feature structures built according to a standardized architecture. The feature structures are sets of feature-value pairs; the value of a feature is either atomic or in itself a feature structure description. The types of values acceptable for a feature are specified in the *signature*, which is organized as a multiple inheritance hierarchy. The following figure shows the gist of the feature architecture of a linguistic sign:

(1)
$$\begin{bmatrix} \textit{sign} \\ \text{PHON } \textit{list(phon-symbol)} \\ \text{SYNSEM} \begin{bmatrix} \text{LOC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD } \textit{head} \\ \text{SPR } \textit{list(synsem)} \\ \text{COMPS } \textit{list(arg)} \end{bmatrix} \\ \text{CONT} \begin{bmatrix} \text{INDEX } \textit{index} \end{bmatrix} \end{bmatrix} \\ \text{NONLOC } \textit{nonloc} \end{bmatrix} \\ \text{RELATIONS } \textit{list(rels)} \\ \text{C-CONT } \textit{c-cont} \end{bmatrix}$$

At the highest level, features whose values can be constrained by selecting heads are collected under the SYNSEM attribute. SYNSEM is divided into local and nonlocal features, nonlocal features carrying information about items that participate in long-distance dependencies. The local feature bundle specifies a range of syntactic and semantic properties of the sign; CAT specifies the part-of-speech specific HEAD features which are propagated by a lexical head to the mother node. It also contains the valence features SPR and COMPS, which contain the specifier and the complements the sign must combine with in order to grow into a well-formed phrase. The value of SPR is a list of SYNSEM objects, whereas the value of COMPS is a list of objects of the type *argument*. This type specifies the SYNSEM value of the valent and tracks its realization with the boolean feature REALIZED:

(2) $\begin{bmatrix} \textit{argument} \\ \text{ARGUMENT } \textit{synsem} \\ \text{REALIZED } \textit{bool} \end{bmatrix}$

At the level of the lexical head, all valents start out with a negative REALIZED specification. Once the valent has been realized, its REALIZED feature switches to positive. The complex representation of valents is adopted for COMPS in place of the cancellation approach to valence as it is assumed in the traditional makeup of the framework: under the cancellation view, a valent are represented by a SYNSEM object which disappears from the valence list once the respective element has been realized. However, there are structures which require the SYNSEM value of realized dependents to be accessible at the mother node (Przepiórkowski 1999; Meurers 1999b; Müller 2008a; Bender 2008). In Chinese, we use this extended valence specification for specific types of serial verb constructions, as described in Section 6.4.[1]

The CONT(ENT) attribute specifies semantic features, the main information being an index variable which identifies a referential or situational argument.

Finally, besides SYNSEM, the three top-level features PHON, RELATIONS and C-CONT contain the phonological form, the semantic relations contributed by the elements composing the sign and the semantic contribution of the mother node. By placing these features outside of SYNSEM, we ensure that their values cannot be specified by selecting heads, which enhances a more constrained theory of selection.

Syntactic composition is mainly determined by the following two principles which are assumed to hold for most languages:

- Head Feature Principle (Pollard and Sag 1994, p. 34): the HEAD value of any headed phrase is structure-shared with the HEAD value of the head daughter.
- Valence Principle: in a headed phrase, for each valence feature F, the F value of the mother is determined as follows:
  - If the valence list of F consists of *synsem* objects, its value corresponds to the head daughter's F value minus the SYNSEM values of its sisters.
  - If the valence list of F consists of *argument* objects, its value corresponds to the head daughter's F value, whereby the valents that are realized as sisters get a positive REALIZED value.

Semantically, the relations list of the mother node is the concatenation of the relations of the daughters. Further, the index of the mother is identified with the index of the head daughter, adjunction structures being an exception because the index of the mother is projected from the adjunct daughter.

2.2 The implementation environment

There are two systems which are used for grammar engineering with HPSG: Trale (Meurers et al. 2002; Penn 2004; Müller 2007) and LKB (Copestake 2002). The implementation presented in this paper uses the Trale system. Trale is a Prolog-based grammar development environment that supports both parsing and generation. It comes with the user interface Grale which allows to display different kinds of linguistic descriptions (parse trees, lexical entries, lexical rules, types, macros etc.).

Additionally to the implementation of descriptions formulated using the tools provided by the framework, macros can be defined in order to make the grammar more readable. Just as the types in the signature, macros are generalizations over linguistic objects that can be organized in an inheritance hierarchy; additionally, macros allow for parametrization.

A Trale grammar can be distributed between an arbitrary number of files, different files containing sets or subsets of linguistic generalizations of a certain type. Thus, file sharing by multiple grammars is straightforward, which eases multilingual grammar development since constraints shared by multiple languages can be organized into separate files (Omitted 2013).

3 Previous work

On the one hand, since the 90's, several studies have provided theoretical HPSG analyses of specific phenomena of Chinese. Formal treatments have been proposed for the NP (Gao 1993; Xue and McFetridge 1995; Ng 1999), serial verb constructions (Lipenkova 2009; Müller and Lipenkova 2009) and the well-known *bǎ*-construction (**?**Lipenkova 2011). Besides, two

[1] For the sake of readability, we will use the cancellation notation for structures that do not require this additional information about the relization status of valents.

dissertations, namely Gang (1997) and Gao (2000), provide overall sketches of HPSG grammars for Chinese.

On the other hand, there are two ongoing efforts in grammar development for Chinese, presented in Wang et al. (2009), Yu et al. (2010), and Zhang et al. (2011, 2012).

Wang et al. (2009) and Yu et al. (2010) adopt a data-driven approach with the aim of developing a HPSG parser for Chinese. Starting out with a small set of assumptions about the grammar (sign structure, grammatical principles and schemata), they manually convert a Chinese treebank into an HPSG treebank; the resulting treebank is used for the extraction of a large-scale lexicon of Chinese.

Zhang et al. (2011) and Zhang et al. (2012) use the HPSG framework to combine grammar engineering and treebank compilation. Besides basic clause structures, the grammar covers the structure of NPs and locative phrases, topic constructions, coverbs, resultative verb compounds and simple *bǎ*- and *bèi*-constructions.

Both projects, though being oriented towards a large-scale data-driven grammar implementation, attempt to stay close to the original version of the framework and minimize the use of language-specific postulates. Our grammar aims to complement these efforts and refine some of the analyses by grounding them on findings from recent descriptive and theoretical research.

4 Theoretical and empirical grounding

Our implementation aims at a theoretically adequate analysis of Chinese which is based on research in theoretical linguistics, but can also be adapted for use in NLP applications.

In the last half-century, Chinese linguistics has been driven by three lines of research:

- The descriptive tradition (Chao 1968; Li and Thompson 1981; Zhū 1982, *i.a.*), mainly followed by native linguists, focusses on the description of semantics, pragmatics and discourse structure. Structural considerations often limit themselves to observations about surface order, whereas syntactic relations are treated in a rather permissive, loosely defined fashion.

- The cognitive line of research, starting with a series of papers by James H.-Y. Tai (Tai 1989, 1992, 1993, *i.a.*), seems to be a natural continuation of the descriptive tradition. Concepts of cognitive linguistics often do not impose strict structural constraints and provide a flexibility which allows for rather intuitive explanations of linguistic phenomena.

- The generative line of research, starting with Huang (1982) and continued in Li (1990), Huang (1992), Sybesma (1999) and Huang et al. (2009), *i.a.*, makes heavy use of theory-internal assumptions adopted from generative grammar. One of the drawbacks of this approach for Chinese is that it sometimes uses data for which empirical support is difficult to find.

In our work, we rely to a great part on descriptive research in order to improve the adequacy of the data and the compliance with intuitions of native linguists about aspects of meaning and usage of linguistic structures. Besides, two corpora, the *Lancaster Corpus of Mandarin Chinese* and the *Modern Chinese Corpus* hosted by Beijing University, are used to backup our empirical claims. Analyses in the generative and cognitive traditions are carefully considered against empirical evidence from these sources. In the following exposition, we often use simplified structures for purposes of illustration in order to ease understanding by non-Chinese speakers.

5 The coverage of the grammar

Our grammar contains a syntactic component which specifies linear order and constituency, a lexicon with about 900 lexical items and a number of lexical rules, as well as a set of macros which are used as 'abbreviations' for recurring descriptions of linguistic objects to ease the work of the grammar writer. The grammar is tested against a testsuite of sentences representing different constituent and clause structures of Chinese. Currently, we are testing the grammar against a larger corpus of real-usage examples of the covered phenomena and extending the lexicon and the grammar as new items and structures arise. The phenomena that can be analyzed at present are:

- NP structure:
  - Internal structure, combination with determiners, numerals and classifiers

  - Prenominal modification: adjectival and possessive modifiers, relative clauses with subject, object and adjunct extraction

- Relative clauses
- Morphological variation: compounding, reduplication, affixation
- Basic clause structures and valence alternations: transitive, intransitive and ditransitive frames; *bǎ-* and *bèi*-construction; serial verb constructions; topic structures; unmarked passives; existential constructions
- Syntactic marking: nominal and verbal *de*-adjunction; verbal *de*-complementation
- Mood and aspect marking
- Modal verbs
- Locative and temporal adjuncts; linear orders of adjuncts
- Resultative and directional constructions

6 Example analyses

This section briefly describes some analyses adopted in the grammar. After describing the set of immediate schemata that we use for Chinese, we consider localizers and locative phrases, existential constructions with locative inversion, aspect marking and serial verb constructions. It should be kept in mind that HPSG works with recursive feature structures which can grow into very detailed and voluminous representations; in the following, we only provide partial descriptions, focussing mainly on the valence and category features as well as features that guide semantic composition. For the sake of readability, we often do not provide full feature path specifications; this has no impact on the theoretical analysis since the omitted feature paths can always be reconstructed using the feature specifications in the signature.

6.1 Set of immediate dominance schemata

We assume binary branching and use immediate dominance (ID) schemata for the combination of heads with complements, specifiers and adjuncts. Adjuncts and complements can combine with heads via two instances of the respective schemata which allow to differentiate between head-initial and non-head-initial phrases; two boolean head features responsible for word order – INITIAL for heads in head-argument structures and PREMODIFIER for modifiers in head-adjunct structures – determine which structure applies in a given phrase instance. Since specifiers always precede their head, only one schema is required for specifier-head combination. Complements and specifiers are selected by their heads, whereas adjuncts select their head. Additionally to these schemata which are common for analyses of different languages, we assume a language-specific ID schema for serial verb constructions. This additional assumption can be justified by the fact that serial verbs occur in languages of limited geographic areas which independently exhibit common structural characteristics (Seuren 1990).

6.2 Localizer phrases and locative PP adjuncts

'Localizers' are particles that specify the position of a figure relative to its ground. In most languages, this semantic relation is expressed by locative prepositions. Chinese has only one generic preposition for signaling the stative position of an entity relative to another entity, namely *zài*; this preposition basically indicates the proximity of two entities without providing more information about the nature of the locative relation. Further specification is required in most cases; in general, only proper names referring to geographic locations (names of cities, countries etc.) can combine with *zài* without additional lexical material that provides further information about the position:

(3) a. Tā zài Běijīng gōngzuò.
he in Beijing work
'He is working in Beijing.'

b. Tā zài wū-*(lǐ) gōngzuò.
he in room-inside.LOC work
'He is working in the room.'

Chinese thus has a small set of postnominal particles (*lǐ* ('inside'), *xià* ('under'), *pángbiān* ('at the side') etc.) which have to be used for further specification of the relative position of the figure.

We analyze localizers as heads selecting for NPs; their semantic index is of sort *locative-rel*. Figure 4 shows the combination of the localizer and the noun for the phrase *wū-lǐ* ('in the room') as used in (3b).

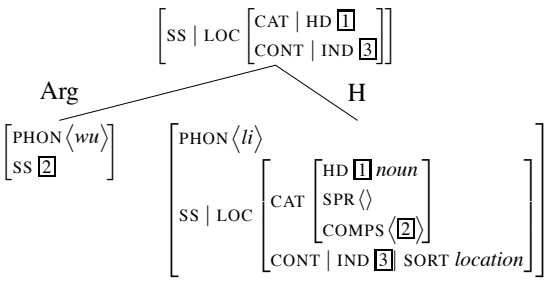
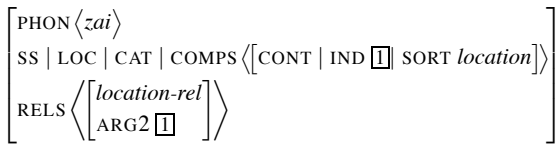


Figure 1: Structure of *wū-lǐ* ('in the room')

The resulting structure can be used in two contexts: on the one hand, they can be used as prepositional complements in locative adjuncts, as illustrated in (3b). On the other hand, they can act as subjects in presentational or existential constructions. Since these positions are prototypical NP positions, we refer to structures composed of an NP and a localizer particle as 'locative NPs'. In the following, we consider the latter usage; locative subjects will be analyzed in Section 6.5.

In locative adjuncts, locative NPs are selected by the generic locative preposition *zài* which signals the static proximity between figure and ground. The semantic combination is regulated by the following constraint in the lexical entry of *zài*:



The direct combination of *zài* with names of geographical locations (3a) is ensured by specifying the indices of these names for the sort *location*.

The position of locative adjuncts is fixed to the position between subject and verb:

(4) a. Tā zài wū-lǐ kànjiàn le guǐ.
he in room-LOC see ASP ghost
'He saw a ghost in the room.'

b. * Zài wū-lǐ tā kànjiàn le guǐ.
in room-LOC he see ASP ghost
'He saw a ghost in the room.'

c. * Tā kànjiàn le guǐ zài wū-lǐ.
he see ASP ghost in room-LOC
'He saw a ghost in the room.'

In order to constrain the possible surface positions of adjunts, they are specified for the boolean head feature PRE-MODIFIER; if the value is positive, the adjunct has to precede the head. The locative preposition *zài*, along with other prepositions heading adjunct PPs, has a positive PRE-MODIFIER value and modifies a VP, that is a verbal projection with a single element in the SPR list and an empty COMPS list.

6.3 Locative inversion

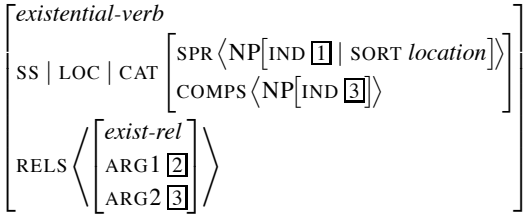
Locative inversion is used to indicate the presence or existence of some entity at a location; the NP denoting the location appears in the specifier position, whereas the entity whose existence is asserted instantiates the complement slot of the existential verb:

(5) Běijīng yǒu hěn duō chē.
Beijing have very many car
'There are many cars in Beijing.'

The sentence-initial position can be occupied either by the name of a geographical location or by a locative NP; locative PPs are not possible in that position:

(6) (*Zài) Běijīng / jiē-shàng yǒu hěn
in Beijing / street-on.LOC have very
duō chē.
many car
'There are many cars in Beijing.'

Thus, the SPR slot of the verb in an existential construction is constrained to an NP that specifies a location (SORT *location*) by virtue of being the name of a geographical location or containing a localizer particle:



Whereas the verb *yǒu* does not additionally constrain the semantics of its complement, other existential verbs may allow only agentive or non-agentive complements (e. g. *zuò* ('to sit'), *táng* ('to

lie') for agentive complements; *guā* ('to hang') for non-agentive complements). These verbs fall into different subclasses of *existential-verb*, the semantic constraints being formulated via selectional restrictions on the index of the complement NP.

6.4 Aspect marking

Chinese has three postverbal aspect markers, as illustrated in the following example:

(7) Tā kàn le / zhe / guo shū.
he read PFV / PROG / EXP book
'He read / is reading / once read the book.'

These markers mark the perfective, durative and experiential aspect, respectively. Naturally, they differ in the range of semantic classes of verbs with which they combine. Their syntactic distribution is identical: they immediately follow the verb. Additional surface material between verb and aspect marker is unacceptable.

In our grammar, aspect marking is analyzed by lexical rules. The rules take a verb as input and output a verb followed by an aspect marker. The relations list of the output verb is the result of appending the aspectual relation contributed by the aspect marker to the relations list of the input verb:

$$\begin{bmatrix} \text{PHON } \boxed{1} \\ \text{SS} \mid \text{LOC} \begin{bmatrix} \text{CAT} \mid \text{HEAD } \textit{verb} \\ \text{CONT} \mid \text{IND } \boxed{3} \end{bmatrix} \\ \text{RELS } \boxed{2} \end{bmatrix} \rightarrow \begin{bmatrix} \text{PHON } \boxed{1} \oplus \langle \textit{le} \rangle \\ \text{RELS} \left\langle \begin{bmatrix} \textit{perfective-rel} \\ \text{ARG } \boxed{3} \end{bmatrix} \right\rangle \oplus \boxed{2} \end{bmatrix}$$

Figure 2: Lexical rule for aspect marking

The lexical rule description specifies only output features which differ from the input. For instance, phonological material is added. The PHON value of the output lexical item is the concatenation of the PHON value of the input and the phonological material associated with the aspect marker. The RELS list starts with an additional *aspect-relation* which takes as argument the event index of the verb. In the example above, the aspectual relation is *perfective-rel*. Further lexical rules are posited for the durative and experiential aspect markers.

6.5 Serial verb constructions

In the basic form, the serial verb construction (SVC) resembles unmarked coordination: two VPs are juxtaposed without overt marking of the relation between them:

(8) Zhāngsān qù chéngshì zhōngxīn mǎi yīfu.
Zhangsan go city center buy clothes
'Zhangsan goes to the city center and buys clothes.'

Depending on the ways in which the two described events can be related by virtue of our world knowledge, different semantic relations can be established between the two VPs. Thus, in (8), the relation would most probably be interpreted as one of *purpose*: Zhangsan goes to the city *in order to* buy clothes. Other possible relations are *causative*, *manner-or-instrument* and *consecutive*. As can be seen in (8), the structure of the SVC may completely underspecify the relation between the two events.

SVCs occur in languages of delimited geographic areas which also share other important structural properties (Seuren 1990). The cross-linguistic occurrence of SVCs justifies the assumption of an additional ID schema, illustrated in Figure 3. The SVC is an instance of a non-headed structure which combines two non-head daughters. The first non-head daughter is a saturated VP; this can be followed from the specification of its COMPS list as a list of spirits.[2] The mother node has a non-empty C-CONT feature which specifies the semantic relation between the two events. Specifically, the RELS feature inside of C-CONT accommodates a relation of the type *svc-relation*, which has the subtypes *causative*, *purpose*, *manner-or-instrument* and *consecutive*.

As described in Gang (1997) and Müller and Lipenkova (2009), the semantic relation can be overtly indicated by perfective or durative aspect marking on one of the VPs. For example, marking of the first VP with the durative aspect marker *zhe* enforces a *manner-or-instrument* reading:

(9) Zhāngsān chàng zhe gē tiàowǔ.
Zhangsan sing DUR.ASP song dance
'Zhangsan sings a song while dancing.'

[2] A spirit is a valent that has already been realized, specified as follows:

(i) $\begin{bmatrix} \textit{argument} \\ \text{ARG } \textit{synsem} \\ \text{REALIZED } + \end{bmatrix}$

(cf. Section 2 on the treatment of valence).

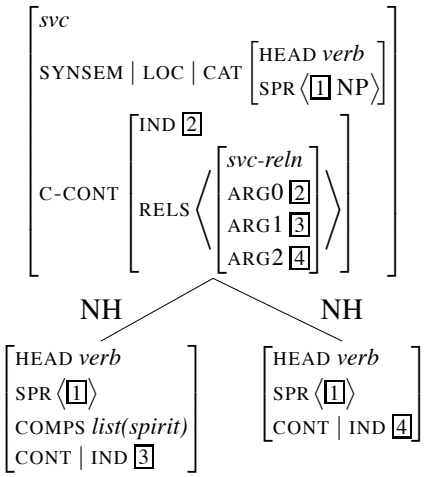


Figure 3: Immediate dominance schema for SVCs

$$\textit{shared-object-svc} \rightarrow \left[\text{NH-DTRS}\left\langle\left[\text{COMPS}\left\langle\left[\begin{array}{l}\text{ARGUMENT } \boxed{1}\\ \text{REALIZED } +\end{array}\right]\right\rangle \oplus \textit{list}\right], \left[\text{COMPS}\left\langle\left[\begin{array}{l}\text{ARGUMENT } \boxed{1}\\ \text{REALIZED } -\end{array}\right]\right\rangle \oplus \textit{list}\right]\right\rangle\right]$$

Figure 4: Valence specification in SVCs with shared objects

This is captured by complex antecedent constraints which relate aspectual relations of the daughters to the relation in the C-CONT feature of the mother. Thus, the following constraint applies for (9):

$$\text{(10)}\quad \left[\begin{array}{l}\textit{svc}\\ \text{NH-DTRS}\left\langle\left[\text{RELS}\left\langle \textit{durative}\right\rangle \oplus \textit{list}\right]\right\rangle\end{array}\right] \rightarrow \left[\text{C-CONT} \mid \text{RELS}\left\langle \textit{manner-or-instrument}\right\rangle\right]$$

A special structural subtype of the SVC is the SVC with a shared object (*shared-obj-svc*): if the objects of the first and the second verb refer to the same referent, the object in the second VP is left unrealized:

(11) Zhāngsān zhǒng cài mài.
Zhangsan plant vegetables sell
'Zhangsan plants vegetables and sells them.'

In this case, the reading is always a purpose reading:

$$\text{(12)}\quad \textit{shared-object-svc} \rightarrow \left[\text{C-CONT} \mid \text{RELS}\left\langle \textit{purpose}\right\rangle\right]$$

In order to establish coreference between the objects in the two VPs and to prevent the realization of the object in the second VP, we make use of the REALIZED feature. Thus, the SYNSEM values of the object valent are identical for both VPs. The valent is realized in the first VP and left unrealized in the second VP (Fig. 4).

For a detailed analysis and formalization of SVCs, the reader is referred to Müller and Lipenkova (2009).

7 Conclusion

In this paper, we have presented our HPSG implementation of a Chinese grammar fragment; after laying out the basic assumptions and concepts of the framework, we have illustrated the use of the formal means provided by the framework for a range of phenomena of Chinese; specifically, we have considered localizers and locative phrases, locative inersion, aspect marking and serial verb constructions. The presented grammar is implemented in the Trale system and is tested against a testsuite which contains both positive and negative examples. Future work includes the extension of the theoretical coverage of the grammar and the systematic use of corpora for the construction of a broader empirical test environment.

References

Bender, Emily M. 2008. Radical Non-Configurationality without Shuffle Operators: An Analysis of Wambaya. In Stefan Müller (ed.), *Proceedings of the 15th International Conference on HPSG, NICT, Keihanna, Japan*, pages 6–24, Stanford, CA: CSLI Publications.

Carpenter, Bob. 1992. *The Logic of Typed Feature Structures*. Tracts in Theoretical Computer Science, Cambridge: Cambridge University Press.

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

Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. Stanford: CSLI Publications.

Copestake, Ann, Flickinger, Daniel P., Pollard, Carl J. and Sag, Ivan A. 2005. Minimal Recursion Semantics: an Introduction. *Research on Language and Computation* 4(3), 281 – 332.

Ding, Picus Sizhi. 2000. A computational study of the *ba* resultative construction: parsing Mandarin *ba*-sentences in HPSG. In *Proceedings of PACLIC 14*, Tokyo, Japan.

Gang, Liu. 1997. *Eine unifikations-basierte Grammatik für das moderne Chinesisch – dargestellt in der HPSG*. Ph. D.thesis, University Constance, SFB 471, FG Sprachwissenschaft, Universität Konstanz, Germany.

Gao, Qian. 1993. Chinese NP Structure. In Andreas Kathol and Carl J. Pollard (eds.), *Papers in Syntax*, OSU Working Papers in Linguistics, No. 42, pages 88–116, Ohio State University: Department of Linguistics.

Gao, Qian. 2000. *Argument Structure, HPSG and Chinese Grammar*. Ph. D.thesis, Ohio State University.

Huang, C.-T. James. 1992. Complex predicates in control. In Richard K. Larson, Sabine Iatridou, Utpal Lahiri and James Higginbotham (eds.), *Control and Grammar*, Dordrecht: Kluwer.

Huang, C.-T. James, Li, Y.-H. Audrey and Li, Yafei. 2009. *The Syntax of Chinese*. Cambridge Syntax Guides, Cambridge, United Kingdom: Cambridge University Press.

Huang, James C.-T. 1982. *Logical relations in Chinese and the theory of grammar*. Ph. D.thesis, MIT, Massachusetts.

Li, Audrey Yen-Hui. 1990. *Order and Constituency in Mandarin Chinese*. Studies in Natural Language and Linguistic Theory, Dordrecht: Kluwer Academic Publishers.

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

Lipenkova, Janna. 2009. *Serienverbkonstruktionen im Chinesischen und ihre Analyse im Rahmen von HPSG*. Masters Thesis, Freie Universitt Berlin.

Lipenkova, Janna. 2011. Reanalysis of obligatory modifiers as complements in the Chinese *bă*-construction. In Stefan Müller (ed.), *Proceedings of the 18th International HPSG Conference*, Stanford: CSLI Publications.

Meurers, Walt Detmar. 1999a. *Lexical Generalizations in the Syntax of German Non-Finite Constructions*. Ph. D.thesis, Eberhard-Karls-Universität, Tübingen.

Meurers, Walt Detmar. 1999b. Raising Spirits (and Assigning Them Case). *Groninger Arbeiten zur Germanistischen Linguistik (GAGL)* 43, 173 – 226.

Meurers, Walt Detmar. 2001. On Expressing Lexical Generalizations in HPSG. *Nordic Journal of Linguistics* 24(2), 161–217.

Meurers, Walt Detmar, Penn, Gerald and Richter, Frank. 2002. A Web-Based Instructional Platform for Constraint-Based Grammar Formalisms and Parsing. In Dragomir Radev and Chris Brew (eds.), *Effective Tools and Methodologies for Teaching NLP and CL*, pages 18 – 25.

Müller, Stefan. 2007. The Grammix CD Rom. A Software Collection for Developing Typed Feature Structure Grammars. In Tracy Holloway King and Emily M. Bender (eds.), *Grammar Engineering across Frameworks 2007*, Studies in Computational Linguistics ONLINE, Stanford: cslip.

Müller, Stefan. 2008a. Depictive Secondary Predicates in German and English. In Christoph Schroeder, Gerd Hentschel and Winfried Boeder (eds.), *Secondary Predicates in Eastern European Languages and Beyond*, Studia Slavica Oldenburgensia, No. 16, pages 255 – 273, Oldenburg: BIS-Verlag.

Müller, Stefan. 2008b. *Head-Driven Phrase Structure Grammar: Eine Einführung*. Stauffenburg Einführungen, No. 17, Tübingen: Stauffenburg Verlag, second edition.

Müller, Stefan. 2013a. The CoreGram Project: A Brief Overview and Motivation. In Denys Duchier and Yannick Parmentier (eds.), *Proceedings of HMGE 2013, Düsseldorf*.

Müller, Stefan. 2013b. The CoreGram Project: Theoretical Linguistics, Theory Development and Verification. Ms. Freie Universität Berlin.

Müller, Stefan and Lipenkova, Janna. 2009. Serial Verb Constructions in Mandarin Chinese. In Stefan Müller (ed.), *Proceedings of the 16th International Conference on HPSG, University of Göttingen, Germany*, Stanford: CSLI.

Müller, Stefan and Lipenkova, Janna. In Preparation. *Mandarin Chinese in Head-Driven Phrase Structure Grammar*. Empirically Oriented Theoretical Morphology and Syntax, Berlin: Language Science Press.

Ng, Say K. 1999. A Double-specifier Account of Chinese NPs Using Head-driven Phrase Structure Grammar. Master thesis, Department of Linguistics, University of Edinburgh.

Penn, Gerald. 2004. Balancing Clarity and Efficiency in Typed Feature Logic Through Delaying. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pages 239–246, Barcelona, Spain.

Pollard, Carl J. and Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. Studies in Contemporary Linguistics, Chicago, London: University of Chicago Press.

Przepiórkowski, Adam. 1999. On Case Assignment and "Adjuncts as Complements". In Gert Webelhuth, Jean-Pierre Koenig and Andreas Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, Studies in Constraint-Based Lexicalism, No. 1, pages 231–245, Stanford, CA: CSLI Publications.

Sag, Ivan A. 1997. English Relative Clause Constructions. *Journal of Linguistics* 33(2), 431–484.

Seuren, P. A. M. 1990. Serial Verb Constructions. In Z. A. M. Joseph Brian D. (ed.), *When Verbs Collide: Papers from the Ohio State Mini-Conference on Serial Verbs*, pages 14 – 32, Ohio: Ohio State University.

Sybesma, Rint. 1999. *The Mandarin VP*. Dordrecht: Kluwer.

Tai, James Y.-H. 1989. Towards a Cognition-based Functional Grammar of Chinese. In *Functionalism and Chinese Grammar*, pages 187 – 226, Chinese Language Teachers Association.

Tai, James Y.-H. 1992. Variation in Classifier Systems Across Chinese Dialects: Towards a Cognition-based Semantic Approach. In *Chinese Language and Linguistics 1: Chinese Dialects*, pages 187 – 226, Taiwan: Academia Sinica.

Tai, James Y.-H. 1993. Iconicity: Motivations in Chinese Grammar. In *Principles and Prediction: The Analysis of Natural Language*, pages 153 – 174, Amsterdam: John Benjamins.

Wang, Xiangli, Iwasawa, Shun'ya, Miyao, Yusuke, Matsuzaki, Takuya, Yu, Kun and ichi Tsujii, Jun. 2009. Design of Chinese HPSG Framework for Data-Driven Parsing. In Olivia Kwong (ed.), *Proceedings of PACLIC 2009*, pages 835 – 842, City University of Hong Kong Press.

Xue, Ping and McFetridge, Paul. 1995. DP structure, HPSG, and the Chinese NP. In *Proceedings of the 14th Annual Conference of Canadian Linguistics Association*, Montreal, Canada.

Yu, Kun, Miyao, Yusuke, Wang, Xiangli, Matsuzaki, Takuya and ichi Tsujii, Jun. 2010. Semi-automatically Developing Chinese HPSG Grammar from the Penn Chinese Treebank for Deep Parsing. In Chu-Ren Huang and Dan Jurafsky (eds.), *Proceedings of COLING 2010*, pages 1417 – 1425, Chinese Information Processing Society of China.

Zhang, Yi, Wang, Rui and Chen, Yu. 2011. Engineering a Deep HPSG for Mandarin Chinese. In *Proceedings of 9th WALR / IJCNLP*, ACL.

Zhang, Yi, Wang, Rui and Chen, Yu. 2012. Joint Grammar and Treebank Development for Mandarin Chinese with HPSG. In Nicoletta Calzolari et al. (ed.), *Proceedings of the LREC'12*, Istanbul, Turkey.

Zhū, Déxī. 1982. *Yǔfǎ jiǎngyì [A manual of grammar]*. Beijing: Shangwu yinshuguan.