Outline of a computerized Chinese grammar enabling English and Swedish translation

Sigurd, Bengt; Gao, Hong

1999

Document Version:
Other version

Lund University, Dept. of Linguistics
Working Papers 47 (1999), 181–199

Outline of a computerized Chinese grammar enabling English and Swedish translation

Bengt Sigurd and Gao Hong

Introduction and abstract

This paper presents a computerized grammar which can analyze and generate a sample of pedagogical Chinese sentences, in particular those mentioned in the textbook *Kinesiska är inte svårt* (‘Chinese is not difficult’) by Göran Malmqvist 1974. Equivalent grammars for English and Swedish have also been constructed allowing translation between the three languages. The grammar model used is the Swetra grammar developed in the MT-project Swetra at the Department of Linguistics, Lund University and used in various translation and generation projects, including the application which generates weather reports in Swedish and English used by SMHI, the Swedish meteorological agency.

The Swetra grammar (see Sigurd (ed.) 1994) is written directly in the DCG (Definite Clause Grammar) format and implemented in Prolog (LPAProlog). It has been used for several languages and adapted to various practical needs. Swetra grammar is characterized by its separate functional (-semantic) representation which is abstract enough to be used as an interlingua in automatic translation. The functional representation of Swetra does not show surface details such as word order, case or agreement. Swetra grammar has ways to account for mode, topic, coordinated clauses and subordinate clauses including relative clauses. The word meanings are represented by a standardized English Machinese according to certain conventions.

The construction of a computerized grammar of Chinese can be based on traditional Chinese grammar but it requires reconsideration and taking a number of decisions. The computerized grammars presented pinpoint the differences between Chinese, English and Swedish. The Chinese, English and Swedish grammars and lexicons are constructed to enable the automatic

translation into or from these languages using the Swetra functional representation as interlingua.

In order to allow writing in both Chinese characters and pinyin the lexical items of the grammar are rended in these two systems. Demo sentences showing the use of the Chinese, English and Swedish grammars are presented.

Chinese lexicon with characters and pinyin

Chinese is normally written by characters, but it may also be written in pinyin, a standardized system using Latin letters and diacritic marks for tones. Pinyin is a kind of phonetic transcription system, that was recommended for general use by the government in 1957. It has, however, not superseded the traditional Chinese writing system. Both writing systems can be handled in our program. The choice only involves the spelling of the lexical items according to a parameter *lex*, that may be set at *c(haracter)* or *p(inyin)*.

There are 4 tones in Mandarin Chinese. If we want to represent the tones there are two systems: diacritic marks or numbers (for details, see Malmqvist 1974). Tone 1 (even) is marked by a line over the vowel (*a™*), the rising tone (tone 2) is marked by an acute accent (*á*); the reversed circumflex is used for the third, falling-rising tone (*a¿*), and the grave accent for the fourth, falling tone (*à*). (In the Prolog program we have to use different markers as the font is restricted, but we will not go inte these problems here.) When numbers are used they are generally placed after the syllable.

The following is a rule showing how the Chinese lexical item meaning ‘busy’ is written. The rule can be rendered in words as follows: There is a Chinese lexical item *(clex)* with the meaning represented by *m(busy,_)* and the category *a(djective)*. It is spelled *[máng]* if the condition (within { } brackets) is met, e.g. if the lexical parameter *(lex)* is set at *p(inyin)*, and ['√¶'], if the lexical parameter is set at *c(haracter)*. The alternatives are given with a semicolon between. In the rule below, *pn* is used for numbered pinyin. Brief comments are given after % as in Prolog programmes.

```
clex(m(busy,_),a) -->   ({lex(p),[máng]}; % pinyin
                        {lex(pn),[mang2]}; % tones marked by numbers
                        {lex(c),['√¶']). % characters
```

Predicative Chinese sentences

The following sentences are found in the first lesson (p. 53) of the textbook mentioned, here given with English and Swedish equivalents.

Chinese (pinyin)	*English*	*Swedish*
wo¿ máng	I am busy	jag är upptagen
wo¿ he¿n máng	I am very busy	jag är mycket upptagen
wo¿ bù máng	I am not busy	jag är inte upptagen
wo¿ bù he¿n máng	I am not very busy	jag är inte mycket upptagen
wo¿ ye¿ he¿n máng	I am also very busy	jag är också mycket upptagen
wo¿ ye¿ bù máng	I am also not/neither busy	jag är inte heller upptagen

The data show that the word order in Chinese is subject (pronoun) first and the predicative adjective last. English and Swedish include a copulative verb not found in Chinese. The copulative finite verb in English and Swedish displays tense (present) – in English also agreement. Three adverbs are included and it is obvious that *he¿n* and its equivalents *very*, *mycket* have to occur immediately in front of the adjective head. We take *he¿n* and its equivalents to be adjective quantifiers in an adjective phrase.

The Chinese data indicate that *ye¿* 'also' should occur before the negation *bù*. In English *also not* is sometimes better rendered by *neither* (and fronted as in *Neither am I busy*). The Swedish translation indicates that the order is *heller inte*, where *heller* is the equivalent of *också* in negative sentences. The order *inte heller* is also acceptable in Swedish.

Following the Swetra phrase structure format, we use rewriting rules where the arrow shows how the functional representation within square brackets to the left of the arrow is rendered by a series of grammatical categories which, eventually, are rendered by words to the right of the arrow. The following is a preliminary rule which covers these data.

```
cs(d,N,[subj(N),pred([P,T]),advl(A1),advl(A2)]) -->
    cnp(N),cadvp1(A1),cadvp2(A2),cap(P). % adjective phrase as predicate
```

This syntactic rule for predicative sentences states that there is a Chinese sentence *(cs)* type which has the functional representation shown within square brackets to the left of the arrow, i.e. a subject with the value represented by the variable *N* given by the first noun phrase, *cnp(N)*, a predicate with the value *P* associated with the adjective phrase *(cap)* and two adverbials *(A1,A2)*, both found in the adverbial phrases *(cadvp1, cadvp2)* before the adjective phrase. Although tense is not marked in Chinese, we have reserved a slot *(T)* for it in the predicate *[P,T]* in order to make the functional representation equivalent to the one needed in English and Swedish, where

tense is marked. The value *d* in the first slot after *cs* indicates declarative mode. The second slot after *cs* includes (information about) the topicalized (preposed) constituent, in this case the subject *N*. We will not investigate the importance of this variable in characterizing and translating sentences in this paper. Two types of adverbials are distinguised in this rule *(cadvp1, cadvp2)* numbered according to order. They are both sentence adverbials, but other adverbial phrases occur in Chinese, of course.

The following phrase rules make it possible to analyze and generate the example sentences under discussion. Word meanings are given in the standard Swetra format *m(L,G)*, where *m* denotes meaning, *L* lexical meaning and *G* grammatical meaning *(sg, pl, pres, past, etc.)*.

```
cnp(m(i,sg)) --> ({lex(p)},[wo¿]; {lex(c)},['Œ“']). % cnp pronoun

cadvp1(m(also,_)) --> ({lex(p)},[ye¿]; {lex(c)},['“≤']).
cadvp2(m(nix,_)) --> ({lex(p)},[bù];{lex{c)},['≤ª']).

cap(A) --> ca(A). % single adjective
cap(m([Q,A])) --> caq(Q),ca(A). % adjective with quantifier

ca(m(busy,_)) --> ({lex(p)},[máng];lex(c)},['√¶']).

caq(m(very,_)) --> ({lex(p)},[he¿n];{lex(c)},['ƒ‹']).
```

The following is a printout of an interaction where the grammar is asked to analyze *wo¿ máng*. Sentences have to be written as a list within [ ] with commas between the words in order to be processed by the Prolog program. (A special printing predicate ‘renprint’ may be applied in order to delete the commas and spell the first word with a capital letter in the print out.) The first line asks for values (solutions) of the variables M,T,F for the Chinese sentence according to the grammar rules. The second line shows the solutions for M(ode) and T(opic), and the third line the functional representation according to our little grammar. Numbers such as _6144, _6852, etc. are arbitrary numbers given by the program if no value is specified for a variable (it is uninstantiated). The term advl([]), means no adverbial, not included among the rules; [] means empty.

```
cs(M, T, F,[wo¿,máng], []).
M=d, T=m(i,sg),
F = [subj(m(i, sg)), pred([m(busy, _6144), _6852]), advl([]), advl([]))]
```

The following printout shows the analysis of the slightly more complex Chinese sentence corresponding to ‘I am not very busy’.

```
cs(M, T, F,[wo¿, bù, he¿n,máng], []), print(F), nl,
    F = [subj(m(i, sg)), pred([[m(very, _6186), m(busy, _6144)], _6852]),
    advl([]), advl(m(nix, _6357))]
```

The following is an equivalent English sentence rule with the same functional representation enabling translation.

```
es(d,N,[subj(N),pred([P,T]),advl(A1),advl(A2)]) -->
    enps(Agr,N),eaux(Agr,m(be,T)),eadvp1(A1),eadvp2(A2),eap(_,P).
```

Note that a form of the auxiliary *be* (copula) is needed in English. It is not represented in the predicate in our rule, only its tense. Alternatively one may assume an underlying (latent) verb ‘be’ , which is not realized in the Chinese sentence. In English subject and object *noun phrases* have to be distinguished in order to choose between *I* and *me*, *he* and *him*, etc. English also requires subject finite verb agreement. The agreement between subject and finite verb in English is handled by the variable *Agr* which takes the value *sg1* for the pronoun *I* allowing the selection of *am*, *sg2* for *you*, *sg3* for *he*, etc. For normal verbs only the third person singular in the present tense *sg3* has to be distinguished.

With the proper lexical additions it is possible to demonstrate automatic translation by calling the Chinese and then the English grammar. The Chinese sentence is inserted as S in the call *cs(M,T,F,S,[]).* After finding the functional representation (F) of the Chinese sentence it is given to the English grammar which is called by *es(M,T,F,X,[])*. The mode, topic and functional representation are to be the same when the computer finds X, which is the translation required.

```
cs(M, T, F,[wo¿, bù, he¿n,máng], []), print(F),nl,
es(M,T,F,X,[]),renprint(X).
F = [subj(m(i, sg)), pred([[m(very, _91935), m(busy, _91893)], _]),
    advl([]),advl(m(nix, _92106))],
X = [I ,am, not,very, busy]
```

Questions and coordination

In Chinese many sentences may be made questions by adding *ma?*. The sentence *nı¿ máng ma?* means *Are you busy?* In English the question mark ? is used and the word order has to be inverted. A declarative sentence is normally marked by a final full stop.

A simple declarative or question sentence may be regarded as a sentence without a following coordinated clause. The following coordinated clause may be said to occur instead of a full stop or question mark. Following Swetra

ideas, we assume a tail constituent which takes care of both the mode marking of single sentences and following coordinated sentences. We denote this constituent *coord* indicating one of its functions. It is realized as a full stop in Chinese declarative clauses and by *ma?* in yes/no questions. The meaning of this constituent is represented as mode and registered as *d(eclarative)* or *q(uestion)* in the mode slot *(M)*.

If there is a following coordinated sentence, its functional representation and the conjunction are registered and included in the functional representation of the top sentence as values of *co*. The following rules indicate how the Chinese rules for *ccoord* may be written.

```
ccoord(d,_,[],[.]) --> ([.]). % declarative, no coordination
ccoord(d,_,[],[]) --> ([]). % declarative, no coordination, but no full stop
ccoord(q,_,[],[]) --> [ma,?]. % question
ccoord(M,N,C,F) --> cconj(C),cs(M1,N1,F). % conj and coordination
```

There is often no equivalent to the English conjunction *and* in Chinese, but it is generally realized as a comma in writing.

The following is the extended Chinese rule, which includes the meaning of the surface constituent *coord* as values of *co* in the functional representation and the mode in a special slot.

```
cs(M,N,[subj(N),pred([P,T]),advl(A1),advl(A2),co(M,N,C,F)]) -->
    cnp(N),cadvp1(A1),cadvp2(A2),cap(P),
    ccoord(M,N,C,F).
```

The quivalent English sentence, however, must have inverse word order as is shown by the following rule, where the mode value is *q(uestion)*.

```
es(q,N,[subj(N),pred([P,T]),advl(A1),advl(A2),co(N,q)]) -->
    eaux(Agr,m(be,T)),enps(Agr,N),
    eadvp2(A2),eap(_,P),eadvp(A1),ecoord(q,N,C,F). % adj in question
```

The following is a printout of the analysis and translation of a question.

```
cs(M, T, F, [wo¿, bù, he¿n, máng, ma, ?], []), print(F), nl, es(M, T, F, X, [])
No.1 : M = q, T = m(i, sg),
F = [subj(m(i, sg)), pred([[m(very,_74232), m(busy, _74190)], pres]),
    obj([]), advl([]), advl(m(nix, _74403)), advl([]), co(q,m(i, sg),[],[])],
X = [am, 'I', not, very, busy, ?]
```

The following example shows the functional representation and translation of *Wo3 mang2, ni3 ye3 mang2* into English and Swedish.

[subj(m(i, sg)), pred(m(busy, _43455)), advl([]), advl([]), co(d,m(i, sg), and, [subj(m(you, sg)), pred(m(busy, _39348)), advl(m(also,_)), advl([]), co(d,m(you, sg), [], [])])]

I am busy and you are also busy.
Jag är upptagen och du är också upptagen.

More complex sentences

We will now touch on the analysis of some more complex sentences which is needed in order to take care of sentences with different numbers and types of constituents and different placement of adverbials.

The following pattern fits Chinese sentences with an additional adverbial phrase before the subject and the next pattern takes account of the case with an adverbial phrase after the subject. These two are the favourite positions for general Chinese adverbials. Note that the functional representation is the same for these two cases, but the topic (the value in the second position) is different.

```
cs(d,A3,[subj(N),pred(m(V,T)),obj(O),advl(A1),advl(A2),advl(A3),
    co(d,A3,C,F2)]) -->
    cadvp(A3),cnp(N),cadvp1(A1),cadvp2(A2),cv(m(V,T)),
    cnpo(O),past(T),ccoord(d,A3,C,F2). % advp before subj np

cs(d,N,[subj(N),pred(m(V,T)),obj(O),advl(A1),advl(A2),advl(A3),
    co(d,N,C,F2)]) -->
    cnp(N),cadvp(A3),cadvp1(A1),cadvp2(A2),cv(m(V,T)),
    cnpo(O),past(T),ccoord(d,N,C,F2). % advp after subj np
```

These two patterns treat transitive and intransitive sentences at the same time, the only differences being that the object is lacking with intransitive sentences *(O=[])*, which has to be allowed by a special rule. The following analysis and translation shows the case where the adverbial is before the subject. The differences in adverbial placement between Chinese and English have not been investigated in detail, and we will not discuss how they can be handled here.

Zai4 Zhong1guo2 wo3 you3 che1.
[subj(m(i, sg)), pred(m(have, pres)), obj(m(wagon, sg)), advl([]), advl([]), advl([in, m(china, prop)]), co(d,_, [], [])]
In China I have a wagon.

The constituent *past(T)* in the rule above needs comment. This is a way of handling the equivalents of past tense in other languages. The *T* of the constituent *past* is realized as *le*, if it is *past*. Present tense is not marked in Chinese. Thus, if *T* is *pres* this constituent is not realized, i.e. *T* is realized as [].

Compare: *wo¿ lái (I come), wo¿ lái le (I came).* This result is achieved by the rules:

```
past(past) --> ({lex(p)},[le];{lex(c)},['¡À']).
past(pres) --> [].
```

But tense is a complex matter in Chinese and we will return to it below.

Relative clauses have the same pattern as ordinary sentences, the only differences being that the relativized constituent is missing. The relativized constituent is the head of the noun phrase, to which the relative clause is attributed. It can therefore be percolated and inserted in the functional representation of the relative clause, using the topic slot of the relative clause (see Noun phrases below). We show the rules for the relative clauses with relativized subject and object, respectively.

```
cs(rel,N,[subj(N),pred(m(V,T)),obj(O),advl(A1),advl(A2),advl(A3)])-->
    cadvp(A3),cadvp1(A1),cadvp2(A2),
    cv(m(V,T)),cnpo(O),past(T). % relativized subj

cs(rel,O,[subj(N),pred(m(V,T)),obj(O),advl(A1),advl(A2),advl(A3)])-->
    cnp(N),cadvp(A3),cadvp1(A1),cadvp2(A2),
    cv(m(V,T)),past(T). % relativized object
```

Particles, auxiliaries, tense and aspect

Chinese is famous for its particles, some of which can be associated with tense and aspect markers in other languages.The following data from the textbook *Kinesiska är inte svårt* illustrate this.

Chinese	*English*
wo¿ xie¿	I write
wo¿ bù xie¿	I do not write
wo¿ xie¿ (xìn) le	I wrote (a letter)
wo¿ xie¿zhe ne	I am writing
wo¿ xie¿zhe xìn ne	I am writing a letter
wo¿ zài xie¿ xìn (ne)	I am writing a letter
wo¿ yı¿jı™ng xie¿ xìn le	I have written a letter
wo¿ yào xie¿ xìn	I will write a letter
wo¿ bú huì xie¿ xìn	I can not write a letter
wo¿ ka™ishı¿ xie¿	I have begun to write
wo¿ ka™ishı¿ xie¿ le	I began to write

The first two sentences illustrate present tense which is unmarked in Chinese (and English). The second sentence shows *do*-support, a complication in English grammar caused by the English requirement that a sentence including *not* must include an auxiliary – *do* if no other is found.

The progressive meaning of *xie¿* 'write' may be expresed by *xie¿zhe* with or without the particle *ne*. Progresssive may also be expressed by *zài*, which may be classified as an auxiliary in front of the main verb. The sentence *wo¿ xie¿ xìn le* may be interpreted as past tense, but also as perfect, and if we want to ascertain the perfect interpretation we have to add *yı¿jı™ng*. The *T* value *perf* is thus realized by two markeres (discontinously) in Chinese. This is no problem in Swetra grammar where e.g. particle verbs and circumpositions are handled in this way. This word is often classified as an adverb with the meaning 'already' by Chinese grammarians, but it may also be regarded as an auxiliary. The classification in (traditional Western) grammatical categories is a well known problem in Chinese grammar as is obvious from the discussions of grammarians. The definitions of tense and aspect are also often different for Chinese. Perfect *yı¿jı¿ng +le* is thus often definied as introducing a new situation. Translation is clearly made difficult beacuse of the more or less subtle differences between what is meant as aspect and tense in different languages.

Future meaning may be rendered by adding *yào* before the verb. This word may similarly be regarded as an adverb, particle or auxiliary. The word *huì* may be regarded as the equivalent of *can*, consequently, as an auxiliary. The word *ka™ishı¿* may be regarded as an equivalent of *begin*, and the last sentence shows how the addition of final *le* gives a past tense (or perfect) meaning.

The situation in Chinese may be handled by assuming a number of constituents (particles, adverbs, auxiliaries) at the proper places. The realization rules must ascertain that the constituents take on the proper value (sometimes nothing,[]) according to the meaning. In some cases a semantic value must be realized in several surface positions. The following are some tentative solutions of these problems.

In the rule below, we assume a predicate of the form *[V,A,T]* where *V* is the meaning of the main verb. *A* is aspect which may take on the values *prog(ressive), perf(ective), future (will), ability* (can) or *beginning (begin)* as shown in the examples.

The following rule shows how the semantic representation may be realized or analyzed with an auxiliary *(caux)* and a particle *(ct)* as required.

```
cs(M,N,[subj(N),pred([V,A,T]),obj(O),advl(A1),advl(A2),advl(A3),
    co(M1,N1,C,F2)]) --> cnp(N),cadvp(A3),cadvp1(A1),
    caux(m(A,_)),cadvp2(A2),cv(m(V,_)), cnpo(O),ct(A),
    ccoord(M1,N1,C,F2). % aux +head verb+particle
```

The success of this rule depends on rules which realize the variables *A* and *T* properly, e.g. the following.

```
caux(m(fut,pres)) --> ({lex(p)},[yao4]);{lex(c)},['“™']).
caux(m(prog,pres)) --> ({lex(p)},[zai4];{lex(c)},['ˇ']).
caux(m(prog,pres)) -->({lex(p)},[zheng4zai4];{lex(c)},['ˇ˝ˇ']).
caux(m(begin,pres)) -->({lex(p)},[kai1shi3];{lex(c)},['ø™ º']).
caux(m(want,pres)) --> ({lex(p)},[xiang3];{lex(c)},['œÎ']).
caux(m(can,pres)) --> ({lex(p)},[neng2];{lex(c)},['ƒ‹']).
caux(m(must,pres)) --> ({lex(p)},[dei3];{lex(c)},['µ√']).
ct(perf) --> ({lex(p)},[le];{lex(c)},['¡À']).
ct(prog) --> ({lex(p)},[ne];{lex(c)},['ƒÿ']).
ct(fut) --> [].
ct(can) --> [].
asp(perf) --> ({lex(p)},[yi3jing1];{lex(c)},['“—æ≠']).
```

The following interactions show the result of such rules and the proper lexical items and conditions.

```
cs(M, T, F, [wo¿, zài, xie¿, ne, .], []), print(F), nl
No.1 : M = d, T = m(i, sg), F = [subj(m(i, sg)), pred([write, prog, pres]),
    obj([]), advl([]), advl([]), advl([]), co(d,m(i, sg), [],[])]

cs(M, T, F, [wo¿, yào, xie¿, .], []), print(F), nl
No.1 : M = d, T = m(i, sg), F = [subj(m(i, sg)), pred([write, fut, pres]),
    obj([]), advl([]), advl([]), advl([]), co(d,m(i, sg), _,[])]

cs(M, T, F, [wo¿, néng, xie¿, .], []), print(F), nl
No.1 : M = d, T = m(i, sg), F = [subj(m(i, sg)), pred([write, can, pres]),
    obj([]), advl([]), advl([]), advl([]), co(d,m(i, sg),[],[])]

cs(M, T, F, [wo¿, ka™ishı¿, xie¿, .], []), print(F), nl
No.1 : M = d, T = m(i, sg), F = [subj(m(i, sg)), pred([write, begin, pres]),
    obj([]), advl([]), advl([]), advl([]), co(d,m(i, sg), [],[])]
```

One may also add rules which allow the treatment of nonfinite verb complexes as in *I will begin to write* where *begin to write* is a complex nonfinite verb prase (for detailed treatment of such phrases, see Sigurd 1994). The following interactions show how such English complex verb phrases are translated into Chinese and Swedish.

```
es(M, T, B,['I',want,to,begin,to,write,.] , []),print(B),nl, cs(M, T, B, Y, []),
    print(Y),nl,ss(M,T,B,Z,[]),print(Z),nl.
```

[subj(m(i, sg)), pred([[write, begin], want, pres]), obj([]), advl([]), advl([]), advl([]), co(d,m(i, sg), [],[])]
[wo¿, yànyi, ka™ishı¿, xie¿, .]
[jag, vill, börja, skriva, .]

es(M, T, B,['I',have,begun,to,write,.] , []),print(B),nl, cs(M, T, B, Y, []), print(Y),nl.
[subj(m(i, sg)), pred([[write, begin], perf, pres]), obj([]), advl([]), advl([]), advl([]), co(d,m(i, sg), [],[])]
[wo¿, ka™ishı¿, xie¿zhe, le, .]
[jag, har, börjat, skriva, .]

Noun phrases

A Chinese noun phrase may consist of a single pronoun or a single noun or coordinated nouns with or without attributes. Adjectival attributes are found before the head as are relative clauses (marked by a following *de* which is also used to mark the genitive in Chinese). The genitive noun phrase is placed before the head noun in Chinese, and it can often be identified with a postnominal attributive prepositional phrase (or adverbial prepositional phrase) in other languages. There are no prepositional phrases found after the head in Chinese.

Some of this is illustrated by the following rules.

cnp(m(i,sg)) --> ({lex(c)},['Œ“'];{lex(p)},[wo3]).
cnp(m(he,sg)) --> ({lex(c)},['À°'];{lex(p)},[ta1]).
cnp(m(you,sg)) --> ({lex(c)},['ƒ„'];{lex(p)},[ni3]).
cnp(m(we,pl)) --> ({lex(c)},['Œ“√«'];{lex(p)},[wo3men2]).
cnp(m(who,_)) --> ({lex(c)},['À≠'];{lex(p)},[shei2]).
cnp(m(what,_)) --> ({lex(c)},[' ≤√¥'];{lex(p)}, [shen2me1]).
cnp(m(china,prop)) --> ({lex(c)},['÷–π˙'];{lex(p)},['Zhong1guo2']).

A noun phrase may consist of a noun only or a noun preceded by an adjective phrase.

cnp(N) --> cn(N).
cnp([A,N]) --> cap(A),cn(N).

Chinese belongs to those languages which use so called classifiers in certain noun phrases, in particular noun phrases including numerals. This implies the addition of a class word which varies with the semantic category of the head noun. For persons the classifier may be *wei4* or *ge4*, which has become the most general (default) classifier. For books the classifier is *ben3*, for tables and other objects characterized by a surface, *zhang1*. We may handle this in our grammar by adding a classifier, which is in harmony (agrees) with the head

word as indicated by the following rules. The condition *harm(C,N)* ascertains that the classifier and the head noun are on the list of harmonic pairs, illustrated as follows:

```
harm(m(person,_),wei4).
harm(m(book,_),ben3).
harm(m(table,_),zhang1).
```

These rules can be given a more general form involving semantic classes.

The following rule shows how a special type of noun phrase including a demonstrative, numeral and a classifier can be written.

```
cdnp([D,Nu,N]) --> cdem(D),cnum(Nu),class(C),cn(N),{harm(C,N)}.
```

We have described relative clauses as clauses without a constituent (subject or object) above. In Chinese the relative clause is marked by a following *de* and occurs before the head. This is illustrated by the following rule. We will not go into detail here, but give some examples of relative clauses in the interactions below.

```
cnp([D,N,F]) --> cs(rel,N,F),[de],cnp([D,N]). % cnp med rel clause
```

Some further Chinese lexical items

```
cdem(m(this,_)) --> ({lex(c)},['’,']; {lex(p)},[zhe4]).
cn(m(chinese,prop)) -->
    ({lex(c)},['÷–π˙ª∞'];{lex(p)},['Zhong1guo2hua4']). % Chinese language
cn(m(chineseperson,sg)) -->
    ({lex(c)},['÷–π˙»À'];{lex(p)},['Zhong1guo2ren2'];).
cn(m(wagon,_)) --> ({lex(c)},['≥µ'];{lex(p)},[che1]).
cn(m(money,_)) --> ({lex(c)},['«Æ'];{lex(p)},[qian2]).
cn(m(person,_)) --> ({lex(c)},['»À'];{lex(p)},[ren2]).
cn(m(letter,_)) --> ({lex(p)},[xìn];{lex(c)},['–≈']).
cn(m(book,_)) --> ({lex(p)},[shu™];{lex(c)},['Æ—']).
cn(m(table,_)) --> ({lex(p)},[zhuo1zi];{lex(c)},['◊¿◊”']).
```

Some Chinese verbs

The following are some Chinese verbs occuring in our examples:

```
cv(m(write,_)) --> ({lex(c)},['–¥'];{lex(p)},[xie3]).
cv(m(come,_)) --> ({lex(c)},['¿¥'];{lex(p)},[lai2]).
cv(m(speak,_)) --> ({lex(c)},['Àµ'];{lex(p)},[shuo1]).
cv(m(learn,_)) --> ({lex(c)},['—ß'];{lex(p)},[xue2]).
cv(m(like,_)) --> ({lex(c)},['∞Æ'];{lex(p)},[ai4]).
cv(m(be,_)) --> ({lex(c)},[' «'];{lex(p)},[shi4]). % with nominal object
cv(m(have,_)) --> ({lex(c)},['”–'];{lex(p)},[you3]).
```

Some Chinese adjectives

```
cap(A) --> ca(A).
cap([A,D]) --> cqadv(A),ca(D). % he¿n máng
cqadv(m(very,_)) --> ({lex(c)},['ʃ‹'];{lex(p)},[hen3]).
ca(m(busy,_)) --> ({lex(c)},['√¶'];{lex(p)},[mang2]).
```

In our preliminary grammar we have distinguished 3 types of adverbials based on word order. The first two categories are sentence adverbs. The category *cadvp* is a general category including time, place and manner adverbials, realized as adverbs, prepositional phrases or subjunctional clauses. Chinese prepositional phrases may be realized as postpositional phrases and a better general term may be adpositional phrases. Similarly, subjunctional phrases may be realized with postsubjunctions. We are not going into all these complications in detail here. The following rules illustrate Chinese adverbial phrases.

```
cadvp2(m(nix,_)) --> ({lex(c)},['≤ª'];{lex(p)},[bu4]).
cadvp2(m(neither,_)) --> ({lex(c)},['“≤','≤ª'];{lex(p)},[ye3,bu4]).
cadvp1(m(also,_)) --> ({lex(c)},['“≤'];{lex(p)},[ye3]).
cadvp(P) --> cprepp(P).
cprepp([P,N]) -->cprep(P),cnp(N). % adverbiell prep phrase
cprepp([P,N]) --> cprep(P),cnp(N),cpost(P). % with postposition
cprepp([P,N]) --> cnp(N),cpost(P).
cprep(in) --> ({lex(p)},[zai4];{lex(c)},['‘⁄']).
cpost(in) --> ({lex(p)},[li3];{lex(c)},['¿Ô']).

cadvp([C,F]) --> cprsubj(C),cs(subj,T,F). % adverbial clauses
cadvp([C,F]) --> cs(subj,T,F),cposubj(C).
cadvp([C,F]) --> cprsubj(C),cs(d,T,F),cposubj(C). % with postpos

cposubj(as) --> ({lex(p)},[suo3yi3];{lex(c)},['À˜“‘']).
cprsubj(as) --> ({lex(p)},[yin1wei2];{lex(c)},['“ÚŒ™']).

cposubj(when) --> ({lex(p)},[de,shi2,hou1];{lex(c)},['µƒ ±ʃÚ']).
cprsubj(when) --> ({lex(p)},[zai4];{lex(c)},['‘⁄']).
```

The demo examples below illustrate the different cases.

Fragments of an equivalent English grammar

The English sentence patterns to be presented are constructed to enable the use of the same functional representation and word meaning representations as the Chinese patterns. English patterns have to take account of the preposing of question words and *do*-support as is shown by the following two rules. The first rule shows both fronting and *do*-support, the second *do*-support caused

by *not* represented as *m(nix,_)*. We do not show the variants with different placement of adverbial phrases.

```
es(q,W,[subj(N1),pred(m(V,T)),obj(W),advl(A1),advl(A2),advl(A3),
    co(q,W,C,F2)]) -->
    {(W=m(who,_);W=m(what,_))},enp(_,W),eaux(Agr,m(do,T)),eadvp2(A2)
    ,enps(Agr,N1),evt(m(V,inf)),eadvp(A1),eadvp(A3),  ecoord(q,W,C,F2).  %
    what does he learn

es(d,N1,[subj(N1),pred(m(V,T)),obj(N2),advl(A1),advl(m(nix,_)),
    advl(A3),co(N1,M,C,F2)]) -->
    enps(Agr,N1),eaux(Agr,m(do,T)),eadvp2(m(nix,_)),evt(m(V,inf)),
    enpo(_,N2),eadvp(A1),eadvp(A3),ecoord(d,N1,C,F2). % trans with not

es(d,N1,[subj(N1),pred(P),obj(N2),advl(A1),advl(A2),advl(A3),
    co(M,N1,C,F2)]) --> enps(Agr,N1),eadvp1(A1),evt(Agr,P),
    enpo(_,N2),eadvp2(A2),{A2/=m(nix,_)},eadvp(A3),ecoord(M,N1,C,F2).    %
    Transitive sentence without not
```

The following is the general pattern for sentences with auxiliaries with or without *not*.

```
es(M,N,[subj(N),pred([V,A,T]),obj(O),advl(A1),advl(A2),advl(A3),
    co(M,N,C,F2)]) -->
    enps(Agr,N),eadvp1(A1),({A2=m(nix,_)},eaux(Agr,m(do,T),G),
    eadvp2(A2),ecomplvt(m(V,G));eaux(Agr,m(A,T),G),eadvp2(A2),
    ecomplvt(m(V,G))),enpo(_,O),eadvp(A3),ecoord(M,N,C,F2).
```

English noun phrases

We will only give a few examples of English noun phrases. English has to distinguish between subject *(nps)* and object np *(npo)*, because of the different forms of personal pronouns. We will not discuss complex English noun phrases with relative clauses, nor the special problems of definite and indefinite articles. Noun phrases have to carry an agreement variable (Agr) in order to fit the requirements of the finite verb.

```
enps(sg1,m(i,sg)) --> ['I'].
enps(Agr,N) --> enp(Agr,N).
enpo(sg,m(i,sg)) --> [me].
enpo(_,N) --> enp(Agr,N).
enp(Agr,N) --> eart(Art),en(Agr,N).
enp(Agr,[A,N]) --> eart(Art),ea(_,A),en(Agr,N).
enp(Agr,N) --> en(Agr,N).
en(sg,m(wagon,sg)) --> [wagon].
en(sg,m(mate,sg)) --> [mate].
en(sg,m(person,sg)) --> [person].
en(pl,m(person,pl)) --> [persons].
```

```
en(sg,m(china,prop)) --> ['China'].
en(sg,m(book,sg)) --> [book].
```

Some English verbs

Due to the agreement requirements the English lexicon must include a number of forms for different persons and tenses. These forms can be derived by morphological rules (adding *-s* in the plural etc. We refer the reader to Sigurd 1994 for details).

```
evt(sg1,m(have,pres)) --> [have].
evt(sg,m(speak,pres)) --> [speaks].
evt(sg,m(learn,pres)) --> [learns].
evt(pl,m(like,pres)) --> [like].
evt(sg,m(like,pres)) --> [likes].
evt(_,m(write,past)) --> [wrote].
evt(sg1,m(be,pres)) --> [am].
evt(sg,m(be,pres)) --> [is].
evt(pl,m(be,pres)) --> [are].
evt(sg2,m(be,pres)) --> [are].
eaux(_,m(must,pres),inf) --> [must].
eaux(_,m(fut,pres),inf) --> [will].
eaux(_,m(can,pres),inf) --> [can].
eaux(sg,m(perf,pres),part) --> [has].
eaux(sg1,m(do,pres)) --> [do].
eaux(sg2,m(do,pres)) --> [do].
eaux(sg,m(do,pres)) --> [does].
```

Some English adverbial phrases

```
eadvp1(m(also,_)) --> [also]. % också m.m
eadvp2(m(nix,_)) --> [not].
eadvp2(m(neither,_)) --> [neither].
eadvp([P,N]) --> eprep(P),enpo(_,N). % adverbial prep phrase

eadvp([C,F]) --> esubj(C),es(subj,T,F). % adverbial subj clause
eprep(in) --> [in].
esubj(when) --> [when].
esubj(as) --> [as].

eap([A,D]) --> eqadv(A),ea(D). % very busy
eap(A) --> ea(A).
eqadv(m(very,_)) --> [very].
```

Fragments of an equivalent Swedish grammar

Question words have to be preposed in Swedish as in English. This is shown by the following rule where the word order is also inverted.

```
s(q,W,[subj(N1),pred(m(V,T)),obj(W),advl(A1),advl(A2),advl(A3),
    co(M,W,C,F2)]) -->
    {(W=m(who,_);W=m(what,_))},npo(_,W),vt(m(V,T)),nps(_,N1),
    advp(A1),advp2(A2),advp(A3),coord(q,W,C,F2). % Vad gillar han?

s(d,N1,[subj(N1),pred(P),obj(N2),advl(A1),advl(A2),advl(A3),
    co(M,N1,C,F2)]) -->
    nps(_,N1),advp1(A1),advp2(A2),vt(P),npo(_,N2),{N2/=m(who,_)},
    {N2/=m(what,_)},advp(A3),coord(M,N1,C,F2). % Simple transitive, Jag
    kanske inte gillar boken

s(d,A3,[subj(N1),pred(P),obj(N2),advl(A1),advl(A2),advl(A3),co(d,A3,C,F2)]
    ) -->
    advp(A3),vt(P),nps(_,N1),advp1(A1),advp2(A2),npo(_,N2),
    {N2/=m(who,_)},{N2/=m(what,_)},coord(d,A3,C,F2). % Preposed adv
    and inversion
```

Swedish noun phrases

Swedish noun phrases are characterized by complex agreement involving gender, number and definiteness, which we will not go into here. The agreement contraints are taken care of by the variable *Agr* which is percolated to all members of the noun phrase. Due to the existence of different pronominal forms for subject and object, two types of noun phrases have to be distinguished in Swedish too. As in English they are generally realized in identical noun phrases.

```
nps(agr(sg,_,def),m(i,sg)) --> [jag].
nps(Agr,N) --> np(Agr,N).

np(agr(sg,n,def),m(china,prop)) --> ['Kina'].
np(agr(sg,r,def),m(who,_)) --> [vem].
np(Agr,N) --> n(Agr,N).
np(Agr,[N,F]) --> n(Agr,N),[som],s(rel,N,F). % with relative clause
npo(_,m(i,sg)) --> [mig].
npo(Agr,N) --> np(Agr,N).
n(agr(sg,r,indef),m(chineseperson,sg)) --> [kines].
n(agr(sg,r,indef),m(wagon,sg)) --> [vagn].
n(agr(sg,r,def),m(wagon,sg)) --> [vagnen].
n(agr(sg,n,indef),m(table,sg)) --> [bord].
n(agr(sg,n,def),m(table,sg)) --> [bordet].

ap(Agr,[A,D]) --> qadv(A),a(Agr,D). % very busy
ap(_Agr,A) --> a(Agr,A).
qadv(m(very,_)) --> [mycket].
a(agr(sg,r,indef),m(busy,_)) --> [upptagen].
a(agr(pl,_,_),m(busy,_)) --> [upptagna].
```

Some Swedish verbs

```
vt(m(have,pres)) --> [har].
vt(m(come,pres)) --> [kommer].
vt(m(speak,pres)) --> [talar].
vt(m(learn,pres)) --> [lär].
vt(m(like,pres)) --> [gillar].
vt(m(write,pres)) --> [skriver].
vt(m(write,past)) --> [skrev].
vt(m(be,pres)) --> [är].

aux(m(fut,pres),inf) --> [skall].
aux(m(fut,pres),toinf) --> [kommer].
aux(m(fut,pres),inf) --> [kommer].
aux(m(can,pres),inf) --> [kan].
aux(m(perf,pres),part) --> [har].
aux(m(prog,pres),toinf) --> [håller,på].
aux(m(fut,pres),inf) --> [skall].
aux(m(want,pres),inf) --> [vill].
aux(m(must,pres),inf) --> [måste].
aux(m(begin,pres),toinf) --> [börjar].
```

Some Swedish adverbial phrases

```
advp1(m(also,_)) --> [också].
advp2(m(nix,_)) --> [inte].
advp2(m(neither,_)) --> [inte,heller].
advp([P,N]) --> prep(P),npo(_,N). % adverbial prep phras
advp([C,F]) --> subj(C),s(subj,T,F). % adverbial subj clause
prep(in) --> [i].
subj(when) --> [när].
subj(as) --> [eftersom].
```

Interactions

The following interactions show the interplay of the Chinese, English and Swedish grammars. The examples illustrate different Chinese writing: characters or pinyin (with numeral or diacritic tone marking). Basic grammatical constructions are illustrated: declarative and question sentences, transitive and intransitive verbs, auxiliaries, different tenses and aspects, prepositional phrases and subjunctive clauses, relative clauses, even embedded relative clauses and noun phrases with different classifiers.

1. No fronting of question word in Chinese, diacritic pinyin, *do*-support in English, detailed calls.

```
es(M,T,F,[what,do,you,speak,?],[]), print(F),nl,
s(M,T,F,K,[]), print(K),nl,
cs(M,_,F,X,[]),print(X),nl,nl.
```

F=[subj(m(you, sg)), pred(m(speak, pres)), obj(m(what, _98136)), advl([]), advl([]), advl([]), co(m(you, sg), q,[],[])]
K=[vad, talar, du,?]
X= [nı¿, shuo™, shémme, ?]

2. Evaluative adjective, numbered pinyin.

Zhong1guo2hua4 bu4 nan2.
[subj(m(chinese, prop)), pred(m(difficult, _19653)), advl([]), advl(m(nix, _20025)), advl([]), co(m(chinese, prop), d, [], [])]
Kinesiska är inte svårt.

3. Coordinated sentences, Chinese characters.

Jag skriver kinesiska men jag talar inte kinesiska.
[subj(m(i, sg)), pred(m(write, pres)), obj(m(chinese, prop)), advl([]), advl([]), advl([]), co(m(i, sg), _78852, but, [subj(m(i, sg)), pred(m(speak, pres)), obj(m(chinese, prop)), advl([]), advl(m(nix, _68451)), advl([]), co(d,m(i, sg),[], [])])]
I write Chinese but I do not speak Chinese.
Œ“ –¥ ∫∫”Ô µ´ « Œ“ ≤ª Àµ ∫∫”Ô.

4. Relativized object.

I like the book that the Chinese man likes.
[subj(m(i, sg)), pred(m(like, pres)), obj([m(def, _61713), m(book, sg), [subj(m(chineseperson, sg)), pred(m(like, pres)), obj(m(book, sg)), advl([]), advl([]), advl([])]]), advl([]), advl([]), advl([]), co(d,m(i, sg),[], [])]
Wo3 ai4 Zhong1guo2ren2 ai4 de na4 ben3 shu1.

5. Self-embedded relative clauses.

Den kines som den person som jag gillar gillar är upptagen.
[subj([m(def, _26184), m(chineseperson, sg), [subj([m(def, _13971), m(person, sg), [subj(m(i, sg)), pred(m(like, pres)), obj(m(person, sg)), advl([]), advl([]), advl([])]]), pred(m(like, pres)), obj(m(chineseperson, sg)), advl([]), advl([]), advl([])]]), pred(m(busy, _7332)), advl([]), advl([]), advl([]), co(d,[m(def, _26184), m(chineseperson, sg), [subj([m(def, _13971), m(person, sg), [subj(m(i, sg)), pred(m(like, pres)), obj(m(person, sg)), advl([]), advl([]), advl([])]]), pred(m(like, pres)), obj(m(chineseperson, sg)), advl([]), advl([]), advl([])]], [], [])]
The Chinese man that the person that I like likes is busy.
Wo3 ai4 de na4 ge4 ren2 ai4 de na4 wei4 Zhong1guo2ren2 mang2.

6. Subordinate clause as adverbial, initial in Chinese.

Jag är upptagen när jag talar kinesiska i Kina.
[subj(m(i, sg)), pred(m(busy, _68361)), advl([]), advl([]), advl([when, [subj(m(i, sg)), pred(m(speak, pres)), obj(m(chinese, prop)), advl([]),

advl([]), advl([in, m(china, prop)]), co(m(i, sg), _63000, [], [])]]), co(d,m(i, sg), [], [])]
Wo3 zai4 Zhong1guo2 shuo1 han4yu3 de shi2 hou1 wo3 mang2.

7. Progressive and coordination.

Jag håller på att skriva och jag är upptagen.
[subj(m(i, sg)), pred([write, prog, pres]), obj([]), advl([]), advl([]), advl([]), co(m(i, sg), d, and, [subj(m(i, sg)), pred(m(busy, _7719)), advl([]), advl([]), advl([]), co(d, m(i, sg), [], [])])]
Wo3 zai4 xie3zhe ne, wo3 mang2.

Conclusion

This preliminary outline of Chinese grammar based on Swetra grammar shows that the basic patterns can be represented in a way which allows computer analysis, generation and automatic translation of Chinese with reasonable success. There are, of course, many Chinese ways of expression which have not been covered by our rules. Different word orders, in particular order of adverbials have not been covered – nor studied in any detail. The use of prepositions and postpositions has not been described. The study of the lexicon is very incomplete. The outline presented, however, points to interesting future investigations. The application of the approach to the two domains weather forecasts and stock market will be studied further.

References

Malmqvist, Göran. 1974. *Kinesiska är inte svårt.* Stockholm: Aldus.

Sigurd, Bengt (ed.). 1994. *Computerized grammars for analysis and translation.* Lund: Lund University Press.