Sentence Segmentation and Segment Re-ordering for English to Urdu Machine Translation

Huda Sarfraz and Tahira Naseem
huda.sarfraz@nu.edu.pk, tahira.naseem@nu.edu.pk

Abstract

This paper describes an English sentence segmentation and segment re-ordering scheme developed for the facilitation of an English to Urdu machine translation (MT) system. The machine translation system performs reasonably well with sentences of up to five to eight words but beyond that the translation time increases such that it is no longer usable. A short term solution was developed, in which the English sentences were segmented (into segments translatable by the MT system) then these segments were input to the machine translation system one by one. The machine translation system was then able to translate the individual segments in a reasonable amount of time (the total translation time for all segments being less than that for the complete sentence), and then concatenate them, after re-ordering some segments to make the order more suitable for Urdu, to achieve the complete translation of the original sentence.

1. Introduction

This paper describes an English sentence segmentation and segment re-ordering scheme developed for the facilitation of English to Urdu machine translation (MT) system. The MT system was basically developed for webpage translation. The average sentence length on the Internet was found to be 19-20 words. The MT system was producing good results with sentences of up to 5-8 words but beyond that it was taking so long to produce the translations that it was not usable on the Internet. The reason for this immense decrease in efficiency with increase in sentence length is that due to the existence of recursive productions in the English parsing grammar, the complexity of translating a sentence increases exponentially as the number of phrases and clauses in the source sentence increases.

One way to solve this problem is to break this complex task into smaller tasks wherever possible without significantly affecting the quality of translation. The breakdown should be such that the smaller tasks are mutually as independent as possible. Most chunk parsers [1] use this scheme, where the main idea is to first break down a sentence into a stream of chunk using a chunk grammar, the individual chunks are parsed and then attached together to form the complete parse. Unfortunately to use this scheme, the current parser used by the MT system would have to be re-designed. Some MT systems also exist, for example a English-to-Korean MT system [2, 3], that partition sentences for efficient translation. The English to Korean MT system mentioned analyzes the structure of sentences (or patterns of sentences) to partition them.

Some other schemes for increasing the efficiency of machine translation were also studied, but taking some ideas from the two mentioned earlier, a scheme was developed which could be implemented with minimal intrusion into the current system. This scheme involved two steps, segmentation and segment re-ordering. First the English sentence would be segmented, using the parts-of-speech (POS) of the words, such that each segment was individually translatable by the MT system. After this the segments would be translated individually and the translated segments would be concatenated after some segments had been re-ordered to produce the complete translation of the sentence. The re-ordering was required so that the segments would be arranged in a manner that was more natural for the Urdu language.

2. Methodology

This section describes the steps taken to devise the segmentation and segment re-ordering scheme.

First, as a preliminary analysis exercise, 151 sentences were segmented manually (initially into complete smaller sentences, and later into phrases). Since the requirement for the MT system was that it should be able to translate grammatically correct sentences from English language websites, sample sentences were taken

from news items at the BBC World website (www.bbcworld.com).

Using data from the previous step, some rules were determined according to which the sentences could be segmented. These rules were implemented and were then tested on a new set of 165 sentences taken from the BBC World website (www.bbcworld.com). Since some of the segments were phrases rather than being complete sentences, some changes were made to the MT system such that it produced translations for phrases as well. Previously it could translate only complete sentences and formations that occurred as headings or titles that occurred as noun phrases.

The rules were then tested and fine-tuned till the system kept showing significant improvement.

At this point, results showed that the translated Urdu segments, some phrasal segments in particular, could not be concatenated as they were ordered in English. So, it became necessary to also re-order some segments such that they would be readable when concatenated to form the complete Urdu translated sentence. The scheme for re-ordering the segments was determined by re-ordering some sample segments manually.

3. English Sentence Segmentation

This section describes the English sentence segmentation technique implemented and its results.

3.1. Structure of the Segmentation Rules

The analysis showed that several rules could be written that would define where and how a sentence could be segmented. These rules were made up of 3 constituents: a segment indicator sequence, indicator disqualifiers and a segment connector. These will be described in the following sections. (Note: The examples given here have been constructed to demonstrate the rule structure and may or may not make sense with real data.)

3.1.1. Segment Indicator Sequence: A segment indicator sequence is a sequence of indicators (?) where each indicator is either a word (including its part of speech (POS)), a POS or a punctuation mark. The MT system was designed such that a punctuation mark could be identified by categorizing it either as a word or a POS, so for segmentation purposes it was decided to treat punctuation marks as POSs. So in essence, a segment indicator was then defined as either a word (including its POSs) or a POS.

For example, the following segment indicator sequence tells that segmentation is possible when a comma is followed by an article.

Indicator Sequence: `comma:# * art:#`
Sentence: "A mediator is not needed," an official told the Reuters news agency.
Segment 1: "A mediator is not needed,"
Segment 2: an official told the Reuters news agency.

The indicators in the sequence are separated by an asterisk (*) and their POS is indicated after a colon (:). In the example both the indicators are POSs and in cases like this a hash (#) is placed in place of the POS to tell that the indicator itself is a POS.

3.1.2. Indicator Disqualifiers: An indicator disqualifier is a sequence similar to the indicator sequence described above except that each word and/or POS in the sequence will also have a position with respect to the segmentation point (the segmentation point is the first indicator in an indicator sequence). Each element in the sequence will be referred to as a disqualifier element. If a segmentation point is found using a segment indicator, but a match is also found in the sentence with a corresponding indicator disqualifier, then the segmentation point is no longer valid, and segmentation will not take place at that point. Each segment indicator sequence may have more than one segment disqualifier, any one of which, if it matches, can invalidate the segmentation point identified by it.

For example, the following indicator sequence and disqualifiers are used to indicate that a sentence should be segmented where the word 'as' (with the POS subordinate conjunction) occurs, except if the word 'as' is either preceded by the word 'known' (with the POS verb) or is followed by the article 'a' (at position 1 after the segmentation point 'as') and a noun (at position 2 after the segmentation point).

Indicator Sequence: `as:sub_conj`
Disqualifiers: `known:v-1 | a:art+1 * n:#+2`

Here, both "`known:v-1`" and "`a:art+1 * n:#+2`" are disqualifiers, and each element making up the sequences ("`known:v-1`",

"a:art+1" and "n:#+2") is a disqualifier element.

With this indicator sequence and disqualifiers pair, it is possible to segment the following sentence as shown:

Sentence: The crowd began to disperse as the day ended.
Segment 1: The crowd began to disperse
Segment 2: the day ended.

The following sentence however, will not be segmented because the segmentation indicator 'as:sub_conj' will be disqualified by the disqualifier "a:art+1 * n:#+2".

Sentence: This is more commonly referred to as a frog.

3.1.3. Segment Connector: When a sentence is segmented, there are three ways that the segment indicator sequence can be dealt with; it can either be 1) attached to the first segment, 2) attached to the second segment or 3) act as a connecting point between the 2 segments. The segment connector was defined to deal with these three possibilities. An "F" will be used to indicate the first case, an "S" will be used to define the second case, and for the third case the translation of the segment indicator, such that it will properly connect the translated segments, will be entered.

The first case is illustrated by the following example:

Indicator Sequence: said:v
Connector: F
Sentence: Mr Annan said the mediator would work discreetly
Segment 1: Mr Annan said
Segment 2: the mediator would work discreetly

An example illustrating the second case is as follows:

Indicator Sequence: comma:# * pro:#
Connector: S
Sentence: "We are talking about modest reinforcements," he told reporters at Nato European headquarters in Belgium.
Segment 1: "We are talking about modest reinforcements,"
Segment 2: he told reporters at Nato European headquarters in Belgium.

An example illustrating the third case is as follows:

Indicator Sequence: but:coord_conj
Connector: لیکن
Sentence: A solution was found but it was meant to be used temporarily.
Segment 1: A solution was found
Segment 2: it was meant to be used temporarily.

When these two segments are translated, they can be connected using the translation for the word 'but'.

Finally, the three components, a segment indicator, indicator disqualifiers and a connector, will combine to form a segment rule. The hash (#) will be used to denote non-applicability (when a POS is the indicator or disqualifier element, e.g., v:#) or non-existence (when there is no disqualifier for a rule), and the asterisk (*) will be used to separate indicator sequence elements.

3.2. Segmentation Rules

After manual segmentation of 151 sentences, a preliminary set of rules was constructed. The rules were then repeatedly tested and refined on a new set of 165 sentences, until no more significant improvement was seen. The finalized rules are given in Appendix A.

3.3. Segmentation Algorithm

The basic segmentation algorithm was a recursive one that kept segmenting a sentence according to the segmentation rules until no more segmentation was possible. Since segmentation was being done based on POS tags, for each lexical item in the sentence, the highest probability POS was considered.

4. Segment Re-ordering for Urdu Translation

One very significant observation that was made at the initial stages of this activity was that the average English sentence that the MT system was supposed to translate contained multiple prepositional phrases and this significantly increased the parsing time for the sentence. Some modifications were made to the MT system (as described earlier) so that it would also translate prepositional phrases independently. After this modification, a segmentation rule was also constructed that would separate all the prepositional phrases of a sentence into

independent segments (rule no. 40 in Appendix A).

It was noted that the most used rule was the preposition rule, but an anticipated problem also arose with the use of this rule. When a sentence with multiple prepositions was segmented and the translated segments were then concatenated, it resulted in an Urdu sentence where the prepositional order was very unusual and barely understandable. An example of this, taken from the test data, is as follows, where PS1, PS2 etc refer to prepositional segments, i.e., segments formed due the prepositional indicator rule (rule no. 40 in Appendix A); the prepositional segments are also underlined (Note: all examples that follow are from the test set of 165 sentences, as translated by the MT system):

Sentence: "We are talking about modest reinforcements[PS1]," he told reporters at Nato European headquarters in Belgium.
Segment: We are talking (rule index: 40)
Segment Translation: ہم بات کر رہے ہیں
Segment Translation: *hum bat kar rahe hain*
Prepositional Segment 1: about modest reinforcements (rule index: 7)
Segment Translation: خاکسار رینفورس َمنٹس کے بارے میں
Segment Translation: *khaksar reinforcements ke baray mein*
Segment: he told reporters (rule index: 40)
Segment Translation: اس نے روداد نویسوں کو بتایا
Segment Translation: *is ne rodad naveeson ko bataya*
Prepositional Segment 2: at Nato European headquarters (rule index: 40)
Segment Translation: ناٹو یورپی کے صدردفتر پر
Segment Translation: *nato europi ke sadr dafter per*
Prepositional Segment 3: in Belgium (rule index: -1)
Segment Translation: بیلجیئم میں
Segment Translation: *belgium ne*
Sentence Translation:
ہم بات کر رہے ہیں خاکسار رینفورس َمنٹس کے بارے میں[PS1] اس نے روداد نویسوں کو بتایا ناٹو یورپی کے صدردفتر پر [PS2] بیلجیئم میں
Sentence Translation: *hum bat kar rahe hain khaksar reinforcements ke baray mein*[PS1] *is ne rodad naveeson ko bataya* **:** *nato europi ke sadr dafter per*[PS2] *belgium ne*[PS3]

(Example 1)

The rule index next to each segment indicates the rule which produced that and the next segment. The last segment shows the rule index to be -1 which means that this is the end of the sentence. See Appendix A for the rules. It can be seen from the translation that the prepositional order that is natural for English is very awkward and not understandable in Urdu. To re-arrange the prepositions such that they would be more understandable in Urdu a re-ordering scheme had to be devised for the prepositional segments. This is described in the next section.

To devise a re-ordering scheme, several of the sentences translated by the MT system after segmentation were translated manually to obtain an ideal ordering of the segments, specifically the prepositional segments that were producing the un-natural order of in the translations.

4.1. Segment Re-ordering Observations

Two observations were made as a result of this manual exercise:

4.1.1. Insertion Point of a Preposition (Sequence): It was noticed that a preposition could not be simply concatenated as it was in English. The following example shows an English sentence, its segments, its translation in the present order, and its ideal translation.

Sentence: The UN has warned of a new "man-made catastrophe" in war-torn Darfur[PS1].
Segment: The UN has warned of a new "man-made catastrophe" (rule index: 40)
Segment Translation: اقوام متحدہ نے مصنوعی تباہی کا بتایا ہے
Segment Translation: *aquame mutahida nay masnawi tabahi ka bataya hay*
Prepositional Segment 1: in war-torn Darfur (rule index: -1)
Segment Translation: وار-ٹورن ڈیرفر میں
Segment Translation: *war-torn darfur mein*
Sentence Translation:
اقوام متحدہ نے مصنوعی تباہی کا بتایا ہے وار-ٹورن ڈیرفر میں [PS2]
Sentence Translation: aquame mutahida nay masnawi tabahi ka bataya hay war-torn darfur mein[PS1]
Ideal Sentence Translation:
قوام متحدہ نے وار-ٹورن ڈیرفر میں[PS2] مصنوعی تباہی کا بتایا ہے
Ideal Sentence Translation: *aquame mutahida nay war-torn darfur mein*[PS1] *masnawi tabahi ka bataya hay*

(Example 2)

In the ideal translation, the prepositional segment (in Darfur) has been inserted right after the subject of the previous segment. This pattern was followed throughout the set of 165 sentences that were being used for testing.

4.1.2. Re-ordering of a Preposition Sequence. The second observation made was that whenever there was a sequence of prepositional segments, the most natural way for them to be ordered in Urdu would be the reverse of the order that they had in English, as can be seen in the following example:

Sentence: Hezbollah seized the soldiers during a cross-border raid[PS1] in July[PS2], triggering the recent conflict with Israel[PS3].
Segment: Hezbollah seized the soldiers (rule index: 40)
Segment Translation: ہیزبولاہ نے سپاہیوں کو پکڑا
Segment Translation: *hizbollah ne sipahion ko pakra*
Prepositional Segment 1: during a cross-border raid (rule index: 40)
Segment Translation: کروس-بورڈر حملے کے دوران
Segment Translation: *cross-border humle ke doran*
Prepositional Segment 2: in July (rule index: 41)
Segment Translation: جولائی میں
Segment Translation: *july mein*
Segment: triggering the recent conflict (rule index: 40)
Segment Translation: حالیہ تنازعہ شروع کرتے ہوئے
Segment Translation: *halia tanazeh shuru karte huay*
Prepositional Segment 3: with Israel (rule index: -1)
Segment Translation: اسرائیل کے ساتھ
Segment Translation: i*srael ke sath*
Sentence Translation:
ہیزبولاہ نے سپاہیوں کو پکڑا کروس-بورڈر حملے کے دوران[PS1] جولائی میں[PS2] حالیہ تنازعہ شروع کرتے ہوئے اسرائیل کے ساتھ[PS3]
Sentence Translation: *hizbollah ne sipahion ko pakra cross-border humle ke doran*[PS1] *july mein*[PS2] *halia tanazeh shuru karte huay israel ke sath*[PS3]

Ideal Sentence Translation:
ہیزبولاہ نے جولائی میں[PS2] کروس-بورڈر حملے کے دوران[PS3] سپاہیوں کو پکڑا ، اسرائیل کے ساتھ [PS3] حالیہ تنازعہ شروع کرتے ہوئے

Ideal Sentence Translation: *hizbollah ne july mein*[PS2] *cross-border humle ke doran*[PS1] *sipahion ko pakra israel ke sath*[PS3] *halia tanazeh shuru karte huay*

(Example 3)

4.2. Segment Re-ordering Algorithm

Keeping the two observations described above in mind the following algorithm was devised for re-arranging the prepositional order for better readability in Urdu:

1. If a sequence of prepositional segments is detected, reverse their order.
2. Insert the prepositional segment (sequence) directly after the subject of the previous segment.

The segments of a sentence were processed independently of each other and had no link with each other till the corresponding translated segments were concatenated. Since these segments were simple Unicode strings and had no other information attached to them, therefore finding the subject of the segment was not possible. It could not simply be the first word of the sentence.

Therefore at the time of concatenation of the translated segments, structural information of the segments was needed to build the complete translation, i.e., the noun phrase that fulfilled the role of the grammatical function subject had to be identified in each translated segment. To accomplish this without a major change in the MT system, the Urdu generating grammar was modified such that in every sentence that was generated, an empty slot was inserted for prepositional phrase insertion right after the subject of the sentence. When concatenating and reordering segments, the prepositional segments were inserted into the empty prepositional phrsase slot of the preceding segment.

6. Results

6.1. Segmentation Results

To analyze the segmentation results, the segments produced by applying the rules given in Appendix A to the test set of 165 sentences, were categorized as either good or bad segments. Good segments were those that were produced as had been intended and were translatable by the MT system. Bad segments were those that were produced when a rule was applied where it shouldn't have been and were not translatable by the MT system. Examples of both can be seen in the following sentences.

Sentence: The BBC is not responsible for the content of external internet sites
Good Segment: The BBC is not responsible (rule index: 40)
Good Segment: for the content of external internet sites (rule index: -1)

(Example 4)

Example 4 shows a sentence that has been segmented into 2 segments, both good (translatable by the MT system), and should result in a complete and correct translation. The rule index at the end of each segment indicates the rule that was applied to produce that and the next segment (see Appendix A for the rules).

Sentence: Mr Annan, his spokesman said, "has not only received a green light from the Israelis but they have also given him a contact point".
Good Segment: Mr Annan (rule index: 8)
Good Segment: his spokesman said (rule index: 0)
Bad Segment: has not only received a green light (rule index: 40)
Good Segment: from the Israelis (rule index: 24)
Good Segment: they have also given him a contact point (rule index: -1)

(Example 5)

Example 5 shows a sentence that has been segmented into 4 good and 1 bad segments. The MT system should translate the 4 good segments and produce a partial translation for the sentence, whereas before the segmentation, it would not have been able to produce any translation for the complete sentence in a reasonable amount of time.

A total of 598 segments were produced from the 165 test sentences, out of which 503 were good segments and 95 were bad segments. The segmentation accuracy was as follows:

(No. of Good Segments / Total No. of Segments) * 100 = 84%

During the rule construction, the focus was more on trying to increase the segmentation than on trying to reduce the segmentation errors, because the increase in segmentation was what was actually improving the performance of the MT system. Out of the 165 sentences tested, 102 sentences had no bad segments and so could be translated completely, and the rest could be partially translated. Due to the segmentation, the MT system was able to produce translations for 124 out the 165 test sentences. The quality of the translations ranged from incomplete and erroneous to complete and correct, but for 124 sentences, some translation was obtained, whereas earlier due to the length of the sentences, almost none were being translated.

6.2. Segment Re-ordering Results

To analyze the re-ordering results the translated sentences (after being segmented and re-ordered) were categorized as follows:

6.2.1. Correctly Re-ordered Sentences. These were the sentences where the re-ordering produced an Urdu sentence with prepositions placed in a natural order. 55 out of the 165 test sentences were of this type. Two of these are shown in Examples 5 and 6.

Sentence: "We are talking about modest reinforcements[PS1]," he told reporters at Nato European headquarters in Belgium.
Segment: We are talking (rule index: 40)
Segment Translation: ہم بات کر رہے ہیں
Segment Translation: *hum bat kar rahe hain*
Prepositional Segment 1: about modest reinforcements (rule index: 7)
Segment Translation: خاکسار رینفورس مینٹس کے بارے میں
Segment Translation: *khaksar reinforcements ke baray mein*
Segment: he told reporters (rule index: 40)
Segment Translation: اس نے روداد نویسوں کو بتایا
Segment Translation: *is ne rodad naveeson ko bataya*
Prepositional Segment 2: at Nato European headquarters (rule index: 40)
Segment Translation: نائو یورپی کے صدردفتر پر
Segment Translation: *nato europi ke sadr dafter per*
Prepositional Segment 3: in Belgium (rule index: -1)
Segment Translation: بیلجیئم میں
Segment Translation: *belgium me*
Translation after Re-Ordering:
ہم خاکسار رینفورس مینٹس کے بارے میں[PS2] بات کر رہے ہیں، اس نے بیلجیئم میں[PS3] نائو یورپی کے صدردفتر پر[PS2] روداد نویسوں کو بتایا
Translation after Re-Ordeing: *hum khaksar reinforcements ke baray mein*[PS1] *bat kar rahe hain, is ne belgium me*[PS3] *nato europi ke sadr dafter per*[PS2]

(Example 5)

Sentence: The group has called for an exchange of prisoners[PS1] with Israel[PS2].
Segment: The group has called (rule index: 40)

Segment Translation: گروہ نے ٹیلیفون کیا ہے
Segment Translation: *giroh ne telephone kiya hai*
Prepositional Segment 1: for an exchange of prisoners (rule index: 40)
Segment Translation: قیدیوں کے تبادلے کے لیے
Segment Translation: *qedion ke tabadle ke liye*
Prepositional Segment 2: with Israel (rule index: -1)
Segment Translation: اسرائیل کے ساتھ
Segment Translation: *israel ke sath*
Translation after Re-Ordering:
گروہ نے اسرائیل کے ساتھ[PS2] قیدیوں کے تبادلے کے لیے[PS1] ٹیلیفون کیا ہے
Translation after Re-ordering: *giroh ne israel ke sath*[PS2] *qedion ke tabadle ke liye*[PS1] *telephone kiya hai*

(Example 6)

6.2.2. Sentences with Bad Segments. These were sentences where the re-ordering turned out to be ineffectual because the sentence translation quality was already low due to bad segmentation. 56 out of the 165 sentences were of this type.

6.2.3. Segments not requiring Re-ordering. Since segment re-ordering was only applied in sentences where there were prepositional segments, those sentences that had no prepositional sentences were not re-ordered. 44 out of the 165 sentences were of this type.

6.2.4. Segments not Re-ordered Deliberately. There was a problem with sentences which had an 'and' segment (produced by rule no. 39, see Appendix A) adjacent to a prepositional segment, but no conclusion was reached about how this should be handled. Since this was occurring in only a few cases, it was decided to not re-order the segments when this happened. This resulted in slightly un-understandable translations, whereas if the re-ordering was allowed in such cases, absolutely un-understandable translations were produced. 8 out of the 165 sentences were of this type.

6.2.5. Incorrectly Segmented Sentences. After the sentences of the 4 types described above, only 2 sentences out the 165 remained which were re-ordered incorrectly.

7. Discussion

This short-term solution worked reasonable well, and served the immediate purpose of making the MT system semi-functional when it was used on internet content. There is still some potential to further refine this solution and improve the result, but it is anticipated that even at its best it will fall short of perfectly segmenting sentences and for the long term, another solution will have to be devised.

Areas where this solution can be improved include:

1. a better set of rules after analysis of a larger and more diverse data set, specifically, removal of rules (in terms of both segment indicators and disqualifiers) that are used very infrequently and addition of new rules. New indicators would produce more segments and new disqualifiers will reduce the error rate of the segmentation.
2. a solution can be devised to handle adjacent 'and' and prepositional segments, also probably obtainable after analysis of a larger and more diverse data set.

8. References

[1] Steven Abney. Parsing by Chunks, In: Robert Berwick, Steven Abney and Carol Tenny (eds.), *Principle-Based Parsing,* Kluwer Academic Publishers, Dordrecht. 1991.

[2] Young-Ae Seo, Yoon-Hyung Roh, Ki-Young Lee and Sang-Kyu Park, "CaptionEye/EK: English-to-Korean Caption Translation System Using the Sentence Pattern", *Machine Translation in the Information Age*, Santiago de Compostela, September 2001.

[3] Yoon-Hyung Roh, Young-Ae Seo, Ki-Young Lee and Sung-Kwon Choi. "Long Sentence Partitioning using Structure Analysis for Machine Translation", *Proceedings of the Sixth Natural Language Processing Pacific Rim Symposium*, Hitotsubashi Memorial Hall, National Center of Sciences, Tokyo, Japan, 2001

Appendix A

The finalized segmentation rules are as follows, the POS tags (including punctuation marks) used in the rules are given in Table A.1:

Table A.1: Names of POS symbols used by the MT system

POS symbol	POS name
v	verb
sub_conj	subordinate conjunction
comma	comma
coord_conj	co-ordinate conjunction
aux	auxiliary
modal	modal
art	article
n	noun
pro	pronoun
gen_pro	genitive pronoun
quant	quantifier
adj	adjective
correlative	correlative
p	preposition

```
// rule 0
said:v
comma:#-1 | sub_conj:#+1
F

// rule 1
say:v
#
F

// rule 2
says:v
#
F

// rule 3
comma:# * saying:v
#
یہ کہتے ہوئے کہ ،

// rule 4
that:sub_conj
v:#+1 | or:coord_conj-1 |
and:coord_conj-1 | aux:#+1 |
modal:#+1
کہ

// rule 5
comma:# * art:#
#
S

// rule 6
comma:# * n:#
#
S

// rule 7
comma:# * pro:#
#
S

// rule 8
comma:# * gen_pro:#
#
S

// rule 9
comma:# * quant:#
#
S

// rule 10
comma:# * adj:#
#
S

// rule 11
if:correlative
#
اگر

// rule 12
comma:# * then:sub_conj
#
، پھر

// rule 13
then:sub_conj
#
پھر

// rule 14
comma:# * or:coord_conj *
that:sub_conj
#
، یا یہ کہ

// rule 15
comma:# * and:coord_conj *
that:sub_conj
#
، اور یہ کہ

// rule 16
or:coord_conj * that:sub_conj
#
یا یہ کہ

// rule 17
and:coord_conj * that:sub_conj
#
اور یہ کہ

// rule 18
comma:# * when:sub_conj
#
، جب
```

```
// rule 19
when:sub_conj
#
جب

// rule 20
comma:# * which:sub_conj
v:#+1 | aux:#+1
جو

// rule 21
because:sub_conj
#
کیونکہ

// rule 22
comma:# * but:coord_conj
v:#+2 | aux:#+2
، لیکن

// rule 23
comma:# * or:coord_conj
#
، یا

// rule 24
but:coord_conj
v:#+1 | aux:#+1
لیکن

// rule 25
neither:correlative
aux:#-1
نہ تو

// rule 26
nor:coord_conj
#
نہ

// rule 27
either:correlative
aux:#-1
کوئی بھی

// rule 28
both:correlative
#
دونوں

// rule 29
comma:# * as:sub_conj
known:v-1
، جیسے

// rule 30
as:sub_conj
known:v-1
جیسے

// rule 31
comma:# * in:p * which:sub_conj
#
، جس میں

// rule 32
in:p * which:sub_conj
#
جس میں

// rule 33
comma:# * where:sub_conj
#
، جہاں

// rule 34
where:sub_conj
#
جہاں

// rule 35
comma:# * while:sub_conj
#
، جب

// rule 36
while:sub_conj
#
جب

// rule 37
declared:v
#
F

// rule 38
comma:# * and:coord_conj
#
، اور

// rule 39
and:coord_conj
aux:#+1 | verb:#+1
اور

// rule 40
p:#
of:p+0 | p:#+1 | aux:#-1
S

// rule 41
comma:# * v:#
#
S
```