THE THIRTEEN STEPS

by

R.H.RICHENS

Honi soit qui mal y pense

SYNOPSIS

Step I Input
II Initial comparison with base chunk dictionary
III Further comparisons with base chunk dictionary
IV Cross references
V Null forms
VI Flexional analysis
VII Syntactic analysis
VIII Interactions
IX Semantic Analysis
X Comparison with target chunk dictionary
XI Syntactic analysis
XII Flexional analysis
XIII Output
Example

Dictionary requirements

1. Base Chunk Dictionary. The entries are in the following form:-
   a chunk a FIBR
   b chunk class b θ (noun theme)
   c flexion class c A
   (e.g. Italian)
   d cross reference d
   e interactions e
   f interlingua f u.Mz

There may be several entries per chunk, for example (Italian):-
   a PIANT
   b N
   c A
   d d PIANG + Past Participle
   e e
   f f
Cross references may be to another chunk, or to a chunk and interlingual word, or to a grammatic category such as past participle above. A chunk with more than one semantic role as PLANT above (several alternatives) will be termed a blob. The two semantic roles are joined by +.

2. Base Null Dictionary
In many inflected languages, the uninflected form has a special semantic role. Thus pig, in English, means pig + singular.

The null dictionary is an inventory of semantically significant uninflected forms. It is of the form:-

<table>
<thead>
<tr>
<th>b chunk class</th>
<th>Ns (noun suffix)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c flexion class e.g. (English)</td>
<td>S</td>
</tr>
<tr>
<td>f interlingua</td>
<td>Pz (singular)</td>
</tr>
</tbody>
</table>

There are, potentially, two null dictionaries, one for suffixes, the other for prefixes.

3. Base Chunk Class Sequence Dictionary
The entries in this dictionary consist of (1) sequences of chunk classes, (2) a single chunk class to which the sequence can be reduced and (3) insertions, such as interlingual syntactic indications.

Thus, for English:-

- AN (adjective noun)
- N (noun)
- QQ (qualifications)

4. Base Grammatic Dictionary
This is similar in form to the base chunk sequence dictionary but instead of chunks under a, grammatic categories such as passive or subjunctive are entered and there are no items corresponding to b and c entries.

5. Target Chunk Dictionary
This is similar to the base chunk dictionary. Thus:-

<table>
<thead>
<tr>
<th>a chunk</th>
<th>a PLANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>b chunk class</td>
<td>b N Ø</td>
</tr>
<tr>
<td>c flexion class</td>
<td>c S</td>
</tr>
<tr>
<td>d cross reference (e.g. English)</td>
<td>d</td>
</tr>
</tbody>
</table>
6. Target Chunk Class Sequence Dictionary.
This is similar to the base chunk class sequence dictionary. It
gives single chunk class equivalents to chunk class sequences, word
re-ordering instructions and insertions. Thus, for English:-
- NA (noun adjective)
- QQ (qualifications)
- N (noun)
- 21 (re-ordering)

7. Target Flexion Table
This contains affixes for each combination of determinants. Thus in
Latin:-
- flexion class A
- case genitive
- number for example: plural
- affix ARUM

8. Special instructions, e.g. capital letter rules

Step I: Input
Feed in one sentence, indicating spaces between words and capital
letters. Store words in order, leaving, say four, empty locations between
each.
Apply capital letter instruction proper to the language. In
English, for instance, all words except the first starting with a capital
letter are marked U (untranslatable).
Number the words in reverse alphabetic order.

Step II: Initial Comparison with Base Chunk Dictionary
Match each word against the base chunk dictionary. The dictionary
chunks are taken in reverse alphabetic order and are matched in succession
against word 1 of the sentence (reverse alphabetic order). A match is
accomplished when a dictionary chunk is wholly contained in the base word,
matching from the front. When a match is made, compare same dictionary
chunk with word 2; if this matches also, proceed to word 3,
otherwise return to the first (reverse order) dictionary chunk beginning
with the same letter and proceed as before. If no match is made when
the first letter of the chunks in the dictionary has overshot the
base word, an indication of untranslatability (U) must be recorded.

Thus the proper name Zachary, which may occur initially, is
unlikely to appear in an English base chunk dictionary, nor will any
component of it match. When the first Y-chunk is matched, Zachary
will be indicated as U.

When a match is made, all the items associated with the chunk are
entered opposite the word matched. If there is an unmatched residue
when a match is made, this is transferred to the location following
the word (the secondary position).

Thus, in Italian, the word PIANTE matches against the chunk PIANT
and E is transferred to the succeeding secondary location.

The procedure described continues till all words are matched or,
marked U.

**Step III: Further Comparisons with Base Chunk Dictionary**

The secondary locations are now numbered in reverse alphabetic
order and the matching procedure is repeated for them. Any residues
are transferred to tertiary positions. The number of positions
required (= maximum number of chunks per word) has not been decided.
Five should suffice for pilot trials.

**Step IV: Cross References**

Base chunks with cross reference items attached are numbered in
reverse alphabetic order of the cross reference item. The latter are
now matched against the base chunk dictionary and base grammatic
dictionary as before. Matching is made either between cross reference
chunk and dictionary chunk or between cross reference chunk + inter-
lingual equivalent and dictionary chunk + interlingual equivalent.
Matching procedure is as in Step II. When a match is made, the items
following the dictionary chunk are entered opposite the base chunk
matched.

**Step V: Null Forms**

The first word of the sentence is subjected to the following
operation. If it contains a chunk with a chunk class of form $\chi^0$
theme taking suffix) and no following chunk, the items in the base null dictionary are entered in the following location. If it contains a chunk of chunk class form \( \theta \chi \) (theme taking prefix) and no preceding chunk, its appropriate null items are entered in the preceding location. For chunks of form \( \chi \theta \) (both prefixing and suffixing), null items are added when there is either no succeeding chunk or no preceding chunk or neither. This operation is carried out on each word in turn.

**Step VI: Flexional Analysis**

The first word is subjected to the following operation. If it contains a suffix of chunk class form \( \chi^{TS} \) (transformer suffix), \( \chi \) is entered as the chunk class of the theme. Thus:

<table>
<thead>
<tr>
<th>PROD V(^{\theta} ) (verb theme)</th>
<th>-&gt; N</th>
</tr>
</thead>
<tbody>
<tr>
<td>UZION N(^{TS} ) (noun transformer suffix)</td>
<td></td>
</tr>
</tbody>
</table>

A similar operation is performed for \( \chi^{TP} \) (transformer prefix) chunks.

Should there be no transformer affixes, the chunk classes and flexion classes of its theme and affixes are compared. All entries corresponding to chunk and flexion classes not present for both theme and affix are expunged. Thus, Latin:

<table>
<thead>
<tr>
<th>VIR</th>
<th>N(^{\theta} )2</th>
<th>N(^{\theta} )3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N(^{\theta} )3</td>
<td>V(^{\theta} )2</td>
</tr>
<tr>
<td></td>
<td>V(^{\theta} )2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ES</th>
<th>N(^{\theta} )3</th>
<th>N(^{\theta} )3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N(^{\theta} )5</td>
<td>V(^{\theta} )2</td>
</tr>
<tr>
<td></td>
<td>V(^{\theta} )1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>V(^{\theta} )2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>V(^{\theta} )3</td>
<td></td>
</tr>
</tbody>
</table>

If all alternatives become expunged, the word is marked U. This procedure is then applied to all the following words. Words marked U are now compared with the base chunk dictionary for suffixes only, matching from behind. If a match is made, the items corresponding to its suffix are entered while the theme remains marked U.

**Step VII: Syntactic Analysis**

The succession of chunk classes making up the entire sentence is now compared with the base chunk class sequence dictionary. It may be
necessary in some languages to have two types of chunk class sequence: (1) immediate and (2) remote. The second type, in which the chunk classes of the sequence may be separated by a number of other chunks will not be considered here. The first chunk sequence in the dictionary say N : N^5 (noun theme - noun suffix in Italian) is compared with the last two (more if the dictionary chunk sequence is longer) chunks of the last word. If these do not match, the penultimate two chunks are compared and so on till the beginning of the sentence is reached or until a match is made. When a match is made, the single chunk class equivalent of the chunk class sequence is entered, together with any insertions. Thus, in Italian:-

\[
\begin{array}{ccc}
\text{FIBR} & \text{N} & \emptyset \\
\text{A} & \text{N^5} & \text{> NQ1}
\end{array}
\]

In all future syntactic matching, the equivalent chunk class is used in matching and not the original chunk sequence. After a match is made, the equivalent chunk class and the following chunk are tested for the same match, then the preceding chunk and the equivalent chunk and so on, till the beginning of the sentence is reached.

After the beginning has been reached, the next chunk sequence is matched with the end of the sentence and passed up to the beginning as before, entering new equivalents and insertions as matches are made.

If a chunk has two alternative word classes, the first only is used in matching.

This process continues till a match is made with a chunk dictionary class sequence forming a complete sentence, i.e. of the form \( \alpha N V N \omega \), where \( \alpha \) represents the beginning and \( \omega \) the end of the sentence. If all the dictionary chunk class sequences are compared and no such match is made, it is ascertained if any chunks have more than one chunk class. If this is so, the second chunk class of the first chunk so affected is used for matching and the entire syntactic matching process is repeated. If this does not achieve an ultimate match with \( \alpha \chi \gamma \zeta \omega \) chunk sequence, the next alternative chunk class is brought into play. All possible combinations of alternatives are tried for \( \alpha \chi \gamma \zeta \omega \) Match but if none achieve this, a return is made to the first set of alternatives which is adopted for the future steps.
It should be noted that the chunk classes used in the syntactical matching may be more elaborate than in the flexional analytic stage.

**Step VIII: Interaction**

Any chunk with an e (interaction) item is treated now. It is tested as follows. A comparison is made between the e item and all chunks bonded directly to it. If a match is made, the interaction stands and other alternatives are expunged. Since the interaction may have a cross reference, it is necessary to test this by comparison with the grammatic dictionary as in Step IV. An example of a double interaction in Latin will illustrate the procedure:

```
    b       d       e
  can
  is
  fel
  em
  persecut { (persequ) V^dep
    (Past Participle) V^s   Passive sum
  us
  est       (sum) V
```

The initial cross referencing of Step IV will show that `persecut` is a blob composed of the theme `persequ` and past participle. Past participle however, interacts with the theme sum to generate a cross reference to passive. We thus get:

```
    b       d       e
  persecut { (perseq) V^dep
    (passive) V^s       -    dep
```

Passive, however, has an interaction too, this time with the chunk class V dep (deponent verb) to generate non-passive. The final result is:

```
    b       d       e
  persecut { (perseq) V^dep
    -                        V^s
```

i.e. chases.

Additional base grammatic dictionary comparisons are required as long as the interactions generate cross references. Often the interaction will merely eliminate alternatives and thus give rise directly to the interlingual equivalent.
Step IX: Semantic Analysis

Every chunk with more than one interlingual equivalent is subjected to a semantic test. The form of this is uncertain. It may be possible to compare the ambiguous chunk with all chunks to which it is bonded. The alternative chosen is that with the greatest number of semantic components in common with the chunks bonded to it. It may also be necessary to compare ambiguous chunks with the sum total of semantic elements from the preceding sentences of the same paragraph. The first half of the translation process is now complete. The base passage has been rendered into a series of bonded interlingual equivalents.

Step X: Comparison with Target Chunk Dictionary

The interlingual words are numbered in reverse alphabetic order and then matched, as in Step II, against the target chunk dictionary. When a match is made, all the items a b c e are entered opposite the interlingual word.

There are no residues to consider.

Step XI: Syntactic Analysis

Syntactic analysis proceeds along similar lines to Step VII. U (untranslatable) words are given the same chunk class as in the base language if it exists in the target language, or the nearest class to it. A procedure for determining the nearest class must await the establishment of formal linguistic criteria for cross-correspondence of chunk classes between different languages. Alternatively, if the base chunk class is not present in the target language, a preferred word class or word classes can be inserted. Thus in translating into English, it is most probable that an untranslatable word is a noun.

Step X will have produced a series of target chunk classes. The first chunk class sequence of the dictionary, together with its bond structure, is compared with the final chunk class sequence of the passage and then forward through the sentence till a match is achieved or the beginning of the sentence is reached. When a match is made, the single chunk class equivalent, the re-ordering instructions and any additional insertions are entered.
Thus, for English:

\[
\begin{align*}
N & \quad Q4 & N & \quad Q & 2 & \quad 2 \\
\text{matches} & & N & \quad & \quad ----\rightarrow & N \\
A & \quad Q4 & \text{against} & A & \quad Q & \quad 1 & \quad 1
\end{align*}
\]

i.e. the bonded sequence noun-adjective is reduced to equivalence with a noun and the order is reversed.

The process continues as in Step VII till a match is achieved with a target chunk sequence of form \( \alpha\chi\zeta\omega \)

**Step XII: Flexional Analysis**

The syntactical analysis produces a sequence of target chunks to which, in inflected languages, grammatic categories will be attached. Thus a sequence in Latin such as:-

\[
\begin{align*}
\{ & b, c \\
\{ & \text{can}, \quad N^g, \quad 3S \\
\{ & \text{nom}, \quad N^s \\
\{ & \text{sing}, \quad N^s \\
\{ & \text{agit}, \quad V^g, \quad 1 \\
\{ & \text{perf}, \quad V^s \\
\{ & \text{sing}, \quad V^s \\
\{ & \text{fel}, \quad N^g, \quad 3S \\
\{ & \text{acc}, \quad N^s \\
\{ & \text{sing}, \quad N^s
\end{align*}
\]

might be obtained.

It is now necessary to match each combination of c items + the grammatic determinants against the entries in the target flexion table. We thus get:-

\[
\begin{align*}
\text{can} & \\
\text{is} & \\
\text{agit} & \\
\text{avit} & \\
\text{fel} & \\
\text{em}
\end{align*}
\]

**Step XIII: Output**

The capitalization rules proper to the language are now applied and the chunks are put out in order of the re-ordering instructions entered at Step XI. Chunks marked U are put out in the form of the original word of which they were a part.
EXAMPLE
ENGLISH TO LATIN

I The the
boys boys
had had
a a
dog dog

b c d e f

II the A q³
boy N⁹ S cm.Mz
had V⁹ have H
a A Pz
dog N⁹ S b28.

b c d e f

III the A q³
boy N⁹ S om.Mz
s N⁹ S P
had V⁹ have H
a A Pz
dog N⁹ S b28

b c d e f

V the A q³
boy N⁹ S om.Mz
s N⁹ S P
have V⁹ H
past V⁹ T₁
a A Pa
dog N⁹ S b28

b c d e f

V the A q³
boy N⁹ S om.Mz
s N⁹ S P
have V⁹ H
past V⁹ T₁
a A Pz
dog N⁹ S b28

N⁹ S Pz
11

VI  the  A  q^3  om.Mz
    boy  N^b  P
    s  N^s  H
    have  V^g  T1
    past  V^s  Pz
    a  A  b28
    dog  N^g  Pz

<table>
<thead>
<tr>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>VI</td>
<td>the</td>
<td>A</td>
<td>q^3</td>
<td>om.Mz</td>
</tr>
<tr>
<td></td>
<td>boy</td>
<td>N^b</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td></td>
<td>s</td>
<td>N^s</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td></td>
<td>have</td>
<td>V^g</td>
<td>T1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>past</td>
<td>V^s</td>
<td>Pz</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>A</td>
<td>b28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>dog</td>
<td>N^g</td>
<td>Pz</td>
<td></td>
</tr>
</tbody>
</table>

VII the  A  q^3  om.Mz
      boy  N^b  NQ1
      s   N^s  Q1  P
      have V^g  VQ2  H
      past V^s  Q2  T1
      a   A   NQ3  Pz
      dog N^g  NQ3  b28

<table>
<thead>
<tr>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>VII</td>
<td>the</td>
<td>A</td>
<td>q^3</td>
<td>om.Mz</td>
</tr>
<tr>
<td></td>
<td>boy</td>
<td>N^b</td>
<td>NQ1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>s</td>
<td>N^s</td>
<td>Q1  P</td>
<td></td>
</tr>
<tr>
<td></td>
<td>have</td>
<td>V^g</td>
<td>VQ2  H</td>
<td></td>
</tr>
<tr>
<td></td>
<td>past</td>
<td>V^s</td>
<td>Q2  T1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>A</td>
<td>NQ3</td>
<td>Pz</td>
</tr>
<tr>
<td></td>
<td>dog</td>
<td>N^g</td>
<td>NQ3</td>
<td>b28</td>
</tr>
</tbody>
</table>

the A  NQ1  SA4  q^3  om.Mz
boy N^b  NQ1
s N^s  Q1  P
have V^g  VQ2  C4  H
past V^s  Q2  T1  Pz
a A  NQ3  B4  Pz
dog N^g  NQ3  b28
N^s  Q3  Pz

<table>
<thead>
<tr>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>A</td>
<td>NQ1</td>
<td>SA4</td>
<td>q^3</td>
</tr>
<tr>
<td>boy</td>
<td>N^b</td>
<td>NQ1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>N^s</td>
<td>Q1  P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>have</td>
<td>V^g</td>
<td>VQ2  C4  H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>past</td>
<td>V^s</td>
<td>Q2  T1  Pz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>A</td>
<td>NQ3</td>
<td>B4  Pz</td>
<td></td>
</tr>
<tr>
<td>dog</td>
<td>N^g</td>
<td>NQ3</td>
<td>b28</td>
<td></td>
</tr>
<tr>
<td>N^s</td>
<td>Q3</td>
<td>Pz</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
VIII
IX
X

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>c</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>puer</td>
<td>N^m</td>
<td>Q1 A4</td>
<td>2er</td>
</tr>
<tr>
<td>plural</td>
<td>N^pl</td>
<td>Q1</td>
<td></td>
</tr>
<tr>
<td>hab</td>
<td>V^g</td>
<td>Q2 C4</td>
<td>2hab</td>
</tr>
<tr>
<td>past</td>
<td>V^s</td>
<td>Q2</td>
<td></td>
</tr>
<tr>
<td>dog</td>
<td>N^m</td>
<td>Q3 B4</td>
<td>3</td>
</tr>
<tr>
<td>sing</td>
<td>N^s sing</td>
<td>Q3</td>
<td></td>
</tr>
</tbody>
</table>

XI

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>c</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>puer</td>
<td>N^m</td>
<td>N m pl</td>
<td>1</td>
</tr>
<tr>
<td>plural</td>
<td>N^pl</td>
<td>2</td>
<td>Q1</td>
</tr>
<tr>
<td>hab</td>
<td>V^g</td>
<td>V</td>
<td>1</td>
</tr>
<tr>
<td>past</td>
<td>V^s</td>
<td>2</td>
<td>Q2</td>
</tr>
<tr>
<td>dog</td>
<td>N^m</td>
<td>N m sing</td>
<td>1</td>
</tr>
<tr>
<td>sing</td>
<td>N^s sing</td>
<td>2</td>
<td>Q3</td>
</tr>
</tbody>
</table>

XII

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>c</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>puer</td>
<td>N^m</td>
<td>N m pl</td>
<td>1</td>
</tr>
<tr>
<td>plural</td>
<td>N^pl</td>
<td>2</td>
<td>Q1</td>
</tr>
<tr>
<td>hab</td>
<td>V^g</td>
<td>V</td>
<td>1</td>
</tr>
<tr>
<td>past</td>
<td>V^s</td>
<td>2</td>
<td>Q2</td>
</tr>
<tr>
<td>dog</td>
<td>N^m</td>
<td>N m sing</td>
<td>1</td>
</tr>
<tr>
<td>sing</td>
<td>N^s sing</td>
<td>2</td>
<td>Q3</td>
</tr>
</tbody>
</table>

XIII

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>c</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>puer</td>
<td>N^m</td>
<td>N m pl</td>
<td>1</td>
</tr>
<tr>
<td>plural</td>
<td>N^pl</td>
<td>2</td>
<td>Q1</td>
</tr>
<tr>
<td>hab</td>
<td>V^g</td>
<td>V</td>
<td>1</td>
</tr>
<tr>
<td>past</td>
<td>V^s</td>
<td>2</td>
<td>Q2</td>
</tr>
<tr>
<td>dog</td>
<td>N^m</td>
<td>N m sing</td>
<td>1</td>
</tr>
<tr>
<td>sing</td>
<td>N^s sing</td>
<td>2</td>
<td>Q3</td>
</tr>
</tbody>
</table>

XIII

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>c</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>puer</td>
<td>pueri</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>can</td>
<td>canem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>em</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hab</td>
<td>habuerunt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>uerunt</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>