12
Météo

In this chapter we describe the Météo system developed by the TAUM group in Montreal for translating weather bulletins from English into French. The system was installed in 1976 and has been in daily operation from the following year to the present time. Its success is based primarily on the completeness and accuracy which could be achieved by the restriction to the sublanguage of meteorological forecasts.

12.1 Historical background

Research on natural language processing began at the University of Montreal in 1965 with the creation of CETADOL (Centre d'Études pour le Traitement Automatique des Données Linguistiques 'Centre for Automatic Linguistic Data Processing Studies') under the direction of Guy Rondeau. At this time, the Canadian government introduced its bilingual policy, which required all official documentation to be available in both French and English. The demands on translation services grew considerably and the Canadian National Research Council began sponsorship of MT research. Between 1968 and 1971 the main focus of the centre became machine translation and the group was renamed TAUM (Traduction Automatique de l'Université de Montréal). The group developed a prototype English to French system, adopting the transfer-based approach and written in the 'Q-systems' software developed by Alain Colmerauer. Already, TAUM had realised the advantages of a sublanguage approach for solving many semantic difficulties of translation (cf. section 8.4) and in 1975 it received a contract to develop a system for translating public weather forecasts.

The task was in many respects ideal for MT treatment. The translation of weather reports is both very boring and highly repetitive. Job satisfaction was already low and there was a high turnover of translation staff when the Canadian government decided to make bilingual weather forecasts available throughout the country. TAUM was asked to develop a system for the Translation Bureau of the Canadian Secretary of State. A prototype was demonstrated in 1976, and Météo began full-time operation in May 1977. Since that date, weather forecasts translated by Météo have been transmitted daily from the Canadian Meteorological Center in Dorval (a suburb of Montreal) for use by the press and television networks. In October 1984 a new version Météo-2 was installed for operation on microcomputers. Météo-2 is written in the programming language GramR developed by John Chandioux Consultants Inc. (Chandioux was the principal designer of the original system). The old mainframe version in the Q-systems program is no longer used: Météo-2 has proved to be faster, more reliable and more cost-effective. A further development in 1989 was the installation by the same company of a Météo system for French–English translation of bulletins issued by the Quebec Weather Office.

In this chapter we describe the original Q-systems implementation of the English–French system.

12.2 The translation environment: input, pre-processing and post-editing

The Météo translation program at the Canadian Meteorological Center is embedded in a larger system which receives the reports from the communication network, pre-processes the data for the Météo system, sends any material which Météo cannot translate to human editors, reformats the Météo output and transmits the final version across the network. At present, the system is translating some 37,000 words every day of the year (over 8 million words p.a.) at an accuracy of over 90%.

Weather reports are received in standard formats, as exemplified in Figure 12.1: a coded heading, a statement of the origin of the bulletin, a list of the regions to which the bulletin applies, the forecast itself, and at the end a terminator. This format is rigidly adhered to. The vocabulary of the bulletins is likewise fixed and predictable, being restricted to the set phrases of the headers, place names, and descriptions of meteorological conditions. This means that it is relatively easy to scan the input text for 'unknown words' which are almost invariably caused by transmission errors. These are virtually the only errors which result in reports being sent for revision by human translators.

The first operation involves the identification of translation units and transforms them into a Q-system format, with individual words and punctuation separated by '+' and the beginnings and ends of phrases by '-01-' and '-02-'. In addition, it marks each unit by an identifier so that, if necessary, revisers can quickly locate the original phrase. These identifiers consist of the source, e.g. Toronto, and a running number. For the bulletin in Figure 12.1 the output would be as in Figure 12.2.

FPCN11 CYYZ 311630
FORECASTS FOR ONTARIO ISSUED BY ENVIRONMENT CANADA AT 11.30 AM EST WEDNESDAY MARCH 31ST 1976 FOR TODAY AND THURSDAY .
METRO TORONTO
WINDSOR.
CLOUDY WITH A CHANCE OF SHOWERS TODAY AND THURSDAY.
LOW TONIGHT 4. H1GH THURSDAY 10.
OUTLOOK FOR FRIDAY... SUNNY
END

Figure 12.1 Weather report as received

-01- $(TORONTO,2) + FORECASTS + FOR + ONTARIO + ISSUED + BY + ENVIRONMENT + CANADA + AT + 11 + H + 30 + AM + EST + WEDNESDAY + MARCH + 31ST + 1976 + FOR + TODAY + AND + THURSDAY + . -02-/

-01- $(TORONTO,3) + METRO + TORONTO + , + WINDSOR + . -02-/

-01- $(TORONTO,4) + CLOUDY + WITH + A + CHANCE + OF + SHOWERS + TODAY + AND + THURSDAY + . -02-/

-01- $(TORONTO,5) + LOW + TONIGHT + 4 + . -02-/

-01- $(TORONTO,6) + H1GH + THURSDAY + 10 + . -02-/

-01- $(TORONTO,7) + OUTLOOK + FOR + FRIDAY + = + SUNNY + . -02-/

Figure 12.2 Formatted weather report

As this example shows, pre-processing includes some regularisation, the original *11.30* becomes *11 + H + 30*; it also includes the expansion of abbreviations, for example *kmh* → *kilometres per hour*, *Jan* → *January*, *BC* → *British Columbia* etc.

It is in this form that the report is sent to the Météo program. The completed translation is formatted as precisely as the original. For the example above, the output is given in Figure 12.3.

In this particular example, it will be seen that Météo failed to translate *high Thursday 10* because of an input error: the numeral *1* instead of the capital letter *I* in *HIGH*. The failure is identified automatically by the system and marked for the attention of the post-editors. Human intervention is therefore involved only when the system fails to translate a phrase, and this is generally because of corrupt input. Post-editing is done at terminals using a minimum of simple commands.

12.3 The translation processes

The linguistic data of Météo consist of three bilingual dictionaries for 'idioms', place names and general (meteorological) vocabulary, and three processing modules for the syntactic analysis of English, the syntactic generation of French, and the morphological generation of French.

Although the TAUM group adopted the transfer-based approach for its major MT research, an essentially 'direct' design was developed for the Météo

U

FPCN1 CYYZ 311630

2

PREVISIONS POUR L ONTARIO EMISES PAR ENVIRONNEMENT CANADA A 11 H 30 HNE MERCREDI LE 31 MARS 1976 POUR AUJOURD HUI ET JEUDI.

1

TORONTO ET BANLIEUE

WINDSOR.

0

NUAGEUX AVEC POSSIBILITE D AVERSES AUJOURD HUI ET JEUDI

0

MINIMUM CE SOIR 4.

-0

HIGH THURSDAY 10.

2

APERCU POUR VENDREDI... ENSOLEILLE.

Figure 12.3 Météo ouput

sublanguage system. There was no need for a transfer module, partly because the 'telegraphic' style of English reports is almost the same structurally as that of French reports. Also absent is a module for the morphological analysis of English; not only is the range of vocabulary limited but morphological variation is restricted, e.g. verbs appear only as present or past participles. Consequently, all lexical variants are entered in the dictionary.

Linguistic analysis is somewhat simplified by the restricted syntactic range of meteorological reports: no pronominal reference, no relative clauses, no passives; above all, the phrases are short. On the other hand, the omission of prepositions and articles caused problems, and led the developers to incorporate analysis procedures based on semantic features.

12.3.1 Dictionary look-up

Translation begins with the extraction of lexical data. The Idioms dictionary indicates that, for example, an English phrase such as *blowing snow* is to be translated as *poudrerie*. For Météo, an 'idiom' is any sequence of more than one word which has to be treated as if it were only one word; 'idioms' are thus defined in a purely practical way, and, more significantly, in bilingual contrastive terms.

The Place-names dictionary contains only those names which differ in English and French (*Newfoundland* → *Terre Neuve*), which are phrasal (*Greater Vancouver* → *Vancouver et banlieue*, *Metro Toronto* → *Toronto et banlieue*), or which must include some linguistic information to ensure correct output (e.g. names which are plural, contain a definite article, or are feminine). All other place names can be omitted, as the program treats any 'unknown' item as if it is a proper name and does not attempt translation.

The main dictionary is the General dictionary containing, as already noted, all morphological forms. Every entry for an English word indicates a French equivalent, a grammatical category, semantic features, and target morphological information. For example (1),

(1) `AMOUNT = N((F,MSR),QUANTITE)`

where `N` indicates a noun, `F` feminine gender and `MSR` a measure noun. The grammatical categories are the traditional ones (adjective, adverb, conjunction, determiner, etc.).

For nouns, the morphological information is restricted to indications of gender (`F` or not) and of plurality. No indication of plural endings for French nouns is necessary, since, with all English forms being entered, they are given directly (e.g. *area* → *région*, *areas* → *régions*). Any differences of number between the languages are treated likewise (*skies* → *ciel*). For adjectives, the syntactic information concerns simply whether it should precede a noun in French or not; and the morphological information indicates the modifications needed to produce feminine and plural forms: `F0` (no change), `F1` (regular *-e* ending), `F2` (*-eau* → *-elle*), `P0` (no change), `P1` (regular *-s* ending), `P2` (*-al* → *-aux*).

The syntactic information for adverbs and verbs is minimal. Entries for adverbs indicate whether they may be attached to adjectives, verbs or prepositions; and verbs are marked simply as either transitive or intransitive, with no other subcategorisation.

Semantic features are attached to nouns, adjectives, adverbs and prepositions. Some examples are shown in Table 12.1.

Feature	*Examples*
time zone	*HAE (heures avancées de l'est)*
month	*juillet, novembre*
day	*lundi, vendredi*
measure	*degré, médiocre, supérieur, à peu près, environ*
place	*secteur, comté, avoisinant, partout, au dessus de*
direction	*est, du nord*
time point	*fin, matin, par la suite, avant*
time duration	*matinée, annuel, pour peu de temps, au cours de*
possibility	*risque, bon, faible, peut-être*
met. phenomenon stationary	*humidité, brume, nuages, chaud, dense*
met. phenomenon falling	*neige, pluie, grêle, abondant*
met. phenomenon blowing	*rafale, vent, venteux, fort*

Table 12.1 Examples of features

Since there is no transfer dictionary, all translational variants are listed in the dictionary, differentiated usually by appropriate semantic features (2).

```
(2) ABOUT = P((MESURE),ENVIRON)
    ABOUT = P((TEMPS),VERS)
    AREA = N((LIEU),REGION)
    AREA = N((CONDITION METEO),ZONE)
    CONSIDERABLE =
                 ADJ((CONDITION METEO,TOMBANT),FORT)
    CONSIDERABLE =
              ADJ((CONDITION METEO,STATIONNAIRE),MARQUE)
```

Choice between variants takes place during the final stages of syntactic analysis, as we shall see.

12.3.2 Syntactic analysis

From an examination of weather reports, the TAUM researchers established a classification of all the phrases which occur. They found that just five types of tree structure were needed. The task of the analysis program is thus defined as the discovery of the particular tree for a given input phrase.

The first phrase type (labelled MET0) consists simply of a list of place names, for example (3).

(3) RED RIVER, INTERLAKE, BISSET, BERENS RIVER.

The next three types have a more complex structure captured by the rule formalisms in (4), where round brackets indicate tree structure, curly brackets alternatives, and square brackets optional elements.

(4a) MET1 $\left(\mathtt{C}\left(\left\{\begin{matrix}\mathtt{Adj}\\ \mathtt{GN}\end{matrix}\right\},[\mathtt{Cmod}]\right),[\mathtt{T}],[\mathtt{L}]\right)$

(4b) MET2 $\left(\mathtt{GN}\left(\left\{\begin{matrix}\textit{highs}\\ \textit{lows}\\ \textit{etc.}\end{matrix}\right\}\right),[\mathtt{T}],[\mathtt{L}],\mathtt{GN}\left(\left\{\begin{matrix}\mathtt{Temp}\ \textit{to}\ \mathtt{Temp}\\ \mathtt{Temp}\end{matrix}\right\}\right),[\mathtt{T}],[\mathtt{L}]\right)$

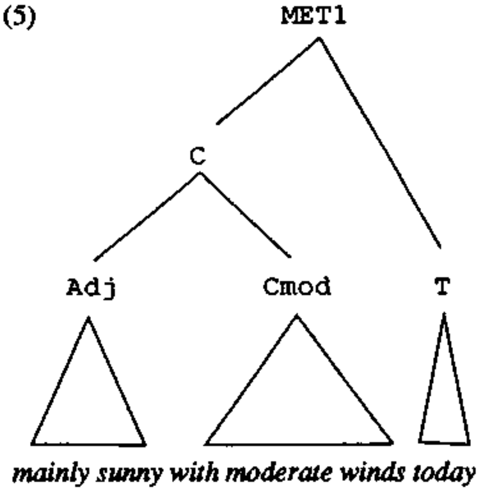
(4c) MET3 $\left(\mathtt{GN}(\textit{outlook for}\ \mathtt{T}),\mathtt{C}\left(\left\{\begin{matrix}\mathtt{Adj}\\ \mathtt{GN}\end{matrix}\right\},[\mathtt{Cmod}]\right),[\mathtt{T}],[\mathtt{L}]\right)$

MET1 (4a) expresses meteorological conditions, e.g. *mainly sunny today, sunny with moderate southwesterly, winds 25 kilometers per hour*. The construction consists of a 'condition' node (C) and an optional time (T) and/or location (L), where the condition is expressed by an adjectival (Adj) or nominal group (GN) followed by an optional complement (Cmod). An example is shown in (5).

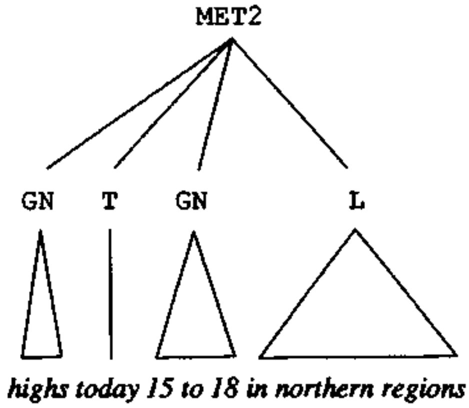
The phrase type MET2 (4b) is for phrases which express maximum and minimum temperatures, e.g. *highs today 15 to 18, lows tonight near 3*. The structure represents an introductory GN such as *highs, lows, temperature range* and so on, followed by a temperature or pair of temperatures. Optional time and place adverbials can appear before or after the temperature(s), as in (6).

Phrase type MET3 (4c) applies to phrases giving the weather outlook, e.g. *outlook for tomorrow, continuing mainly sunny*. The structure is similar to that of MET1, with the addition of a nominal group (GN) before the meteorological condition (C), as exemplified in (7).

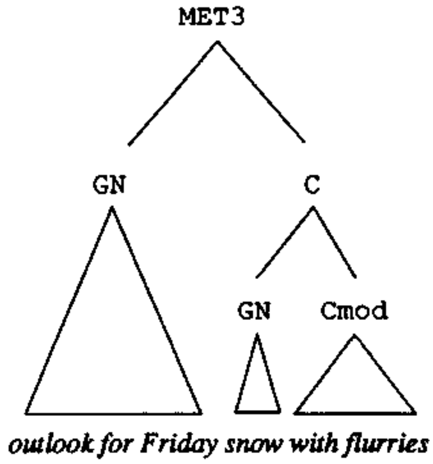
(5)


MET1
C
Adj
Cmod
T
mainly sunny with moderate winds today

(6)


MET2
GN
T
GN
L
highs today 15 to 18 in northern regions

(7)


MET3
GN
C
GN
Cmod
outlook for Friday snow with flurries

The final phrase type covers the bulletin header indicating its origin, for example as in (8).

(8) FORECAST FOR MANITOBA ISSUED BY ENVIRONMENT CANADA AT 6 AM CST APRIL 8TH 1976 FOR TODAY AND FRIDAY.

These are invariant with only place names, dates, days and hours changing. The label for these stereotypical phrases is ORG.

Syntactic analysis is accomplished in three stages by use of a fairly traditional (from the linguistic point of view) bottom-up parsing technique, the computational features of which we shall look at in section 12.4. From the point of view of linguistic processing, it is sufficient to mention that the procedure creates at each stage all possible partial solutions, which are successively pruned as they are combined until, in the final stage, there is just one analysis.

The first stage involves the recognition of dates, hours and degrees of temperature. Strings such as *in the low 20s* are transformed into *21 à 23*, and the English time periods *AM* and *PM* are changed into the French 24-hour equivalents. Next comes the identification and translation of time expressions such as *afterwards*, *this evening*, *on Monday evening*, *early in the evening*, etc. Syntactically they are adverbs or noun phrases optionally preceded by prepositions; semantically they are recognised by semantic 'time' or 'place' features (as in (4) and (5) above). The analysis includes the treatment of conjoined expressions such as *in the afternoon or evening*. In this case an article must also be inserted: *dans l'après-midi ou la soirée.*

The next stage identifies the remaining noun phrases, i.e. those expressing meteorological conditions. The procedure involves both the syntactic categories of lexical items and also their semantic features. Matching these features helps both correct analysis and selection between alternative French translations (it should be recalled that lexical transfer has already taken place, and such cases are treated as ambiguities). For example *heavy* has three possibilities (9) depending on whether it modifies nouns such as *rain*, *wind* and *fog*.

```
(9) ADJ((TOMBANT),ABONDANT)
    ADJ((STATIONNAIRE),DENSE)
    ADJ((SOUFFLANT),FORT)
```

The matching of features ('TOMBANT', etc.) to those attached to the modified nouns ensures the correct renditions: *pluie abondante* 'heavy rain', *vents forts* 'heavy winds', *brouillard dense* 'heavy fog'.

Similar feature matching ensures the translation of *bank* as *rive* or *banc* (10a) and of *around* as *environ* or *vers* (10b), the latter using the definitions given in (2) above.

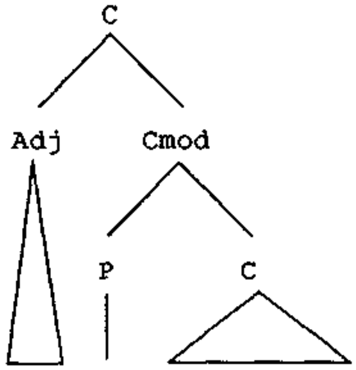
(10a) river banks → *rives du fleuve*
cloud banks → *bancs de nuages*

(10b) around 10 degrees → *environ 10 degrés*
around noon → *vers midi*

The features also help to resolve ambiguous structures such as coordination: compare *gusting winds and snow*, where the adjective modifies only the first noun (because snow does not gust), and *persistent rain and sleet*, where both nouns are included in the scope of the adjective.

Finally, analysis recognises that meteorological conditions (C) can be complements (Cmod) of other conditions (C), thus forming more complex sub-trees (11).

(11)



cloudy with a chance of showers

Furthermore, phrase types (MET1, MET2, etc.) can themselves be conjoined as larger units, 'sentences' (S). The result of syntactic analysis is thus a single tree, dominated by the S node, with English words as terminal nodes and the standardised types of phrase structures specifying the non-terminal nodes.

12.3.3 Syntactic and morphological generation

The task of syntactic generation is to derive the French word order from the structural representation and the information attached to lexical items. Expressions of time and location are placed after expressions of meteorological conditions; adjectives are placed after the nouns they modify (except when indicated otherwise, e.g. *gros*); and articles are inserted. It involves also the selection of the correct prepositions as translations of English *in* with places: *in Montreal*, *in Nova Scotia*, *in Ontario*, *in Manitoba*. It is *à* if the noun never has an article (*à Montréal*), *en* if the noun is feminine (*en Nouvelle-Écosse*), *en* if the noun is masculine and begins with a vowel (*en Ontario*), and *au* if the noun is masculine, begins with a consonant and can have an article (*au Manitoba*).

The final stage of morphological generation is concerned primarily with ensuring the correct endings of adjectives (feminine, plural, etc.) and with contextual adjustments: *le été* → *l'été*, *ce été* → *cet été*, *un beau été* → *un bel été*, *à le* → *au*, *de le* → *du*, etc.

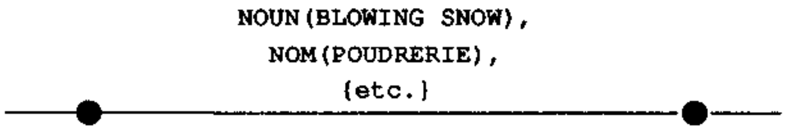
12.4 The computational processes

From a computational point of view, Météo is constructed as a single unified system in the fashion of a **production system** (see section 3.8.6). There is a single data structure and rule-writing formalism for all modules. The data structure is in the form of a **chart** (see section 3.8.5), the arcs of which are labelled with feature bundles. There is no overt control structure to determine the order of application of grammar rules. These are interpreted, in the manner typical of a classical production system, as pairs of pattern matches on the chart and actions which build a new arc on the chart. This production system architecture corresponds to a bottom-up breadth-first parser. The software was written by Alain Colmerauer, and called **Q-systems** ('Q' for 'Quebec', allegedly). Q-systems are regarded as the prototype of the (now well-known) programming language Prolog, which has a similar computational architecture.

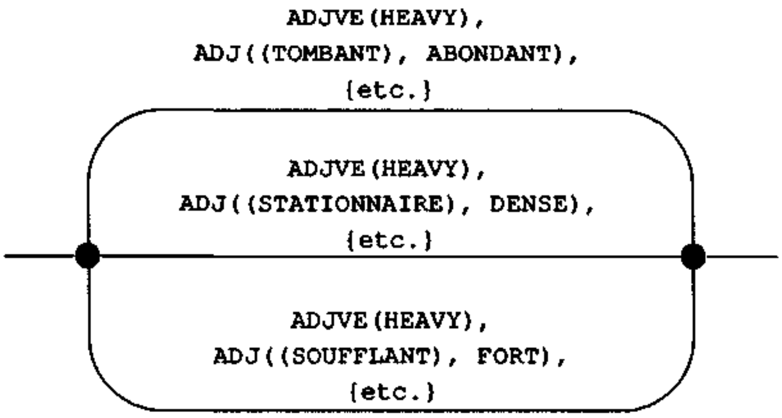
12.4.1 The data structure

As just stated, the data structure in Météo is a chart, the arcs of which are labelled with feature bundles. The initial process of dictionary look-up (section 12.3.1) builds the initial configuration of the chart by constructing one arc for each reading of each lexical item. Thus, multi-word items like *Thunder Bay* or *blowing snow* have a single arc (12a), while an 'ambiguous' word like *heavy* (cf. (9) above) has parallel arcs in the chart (i.e. arcs connecting the same pair of nodes), as in (12b).

(12a)

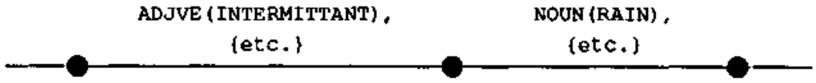


(12b)

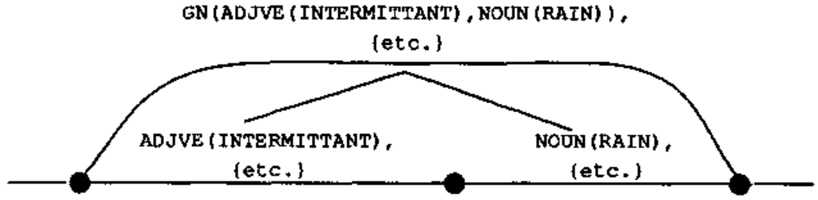


As each rule is applied, the pattern matcher looks for a sequence of arcs as defined by the rule's 'pattern-match' part, and then builds a new arc which spans the whole sequence and which is labelled in accordance with the rule's 'action' part. So, for example, a rule that builds a noun group from the sequence adjective plus noun (e.g. *intermittent rain*) would take the part of the chart in (13a) and build a new arc as in (13b).

(13a)



(13b)



When building a new arc it is essential to indicate which arcs have been subsumed (actually 'used') in its construction: for example, if the adjective is*heavy*, the new arc must indicate which reading of *heavy* has been accepted. It should be noted that the arcs subsumed by rules are not 'thrown away': they are still part of the data structure, and can participate in other rules. In this way, competing alternative partial structures can be maintained until the final complete analysis has built a single arc for the whole structure. A legitimate interpretation is one which has been appropriately labelled by one of the five phrase types given above (METO, MET1, etc.). And likewise, the whole bulletin should conform to the expected structure, i.e. a single header (ORG) followed by any combination of the other phrase types.

From a computational point of view, there is no distinction between the processes of analysis and generation. Like analysis, generation is activated by rules involving pattern matches on the chart and the construction of further arcs. The two processes are, however, kept separate. At the end of analysis, the chart is 'tidied up' and unneeded arcs are removed before the chart is passed to the generation module. But this is essentially a strategy to increase computational efficiency (it is easier to search a smaller database for possible pattern matches), rather than a strict modularisation of the translation procedure as a whole.

12.4.2 Rule formalism

The rules are rather similar to the GETA rules (Chapter 13) in that they define a tree and then describe features on each of the branches of the tree. The left hand side of a rule is a sequence of arcs on a chart and the right hand side specifies the new arc. For example (14) shows the morphological rule which interprets *worse* as the stem *bad* plus a suffix *-er*.

(14) `WORSE == ADJ(BAD,/) + *(ER)`

The rule defines on the left hand side a tree 'WORSE' (a single node) and builds on the right hand side a tree with two branches, one as ADJ with a subordinate element BAD and one as * with a subordinate ER.

Another, more general, rule (15) identifies this as a comparative construction.

(15) `ADJ(A*,/,U*) + *(ER) == ADJ(A*,/,U*,INV(MORE))`

Here `A*` and `U*` are variables: `A*` may apply to any lexical item, and `U*` to any list of features. In the case of the output from (14), `A*` corresponds to `BAD`, `U*` is null (an empty list), and `*(ER)` is an exact match. The result on the right hand side of (15) is a single-node tree which has copied `A*` and `U*` and incorporates the feature `INV(MORE)`.

12.4.3 Rule application

The Q-system procedure is a classical production system in the sense that the rule-set is an unordered set of these linguistic rules. The process is controlled purely by the state of the database (the chart). This determines which of the rules in the rule-set can be applied next. Often, more than one rule can be applied, in which case both possibilities must be tried, theoretically 'in parallel'. However, since this would involve potentially huge computational resources, maintaining perhaps dozens of only slightly different copies of the database, the system in fact applies alternative rules in sequence, and backtracks when appropriate if the first rule applied turns out to be the wrong one. This feature of the Q-systems architecture is now, of course, very familiar to Prolog programmers, but when first implemented in Météo it was an innovation.

There are major disadvantages with this 'unstructured' approach to linguistic processing which are largely avoided in Météo by virtue of its compactness: the rule set is relatively small, and this is because the text type is very restricted. While the production system architecture may be appropriate for Météo, it does not follow that it is generally applicable for translation systems, as indeed the TAUM researchers discovered when they attempted to use the approach in their second much larger MT system, Aviation (see below).

There are potential disadvantages also in the use of a single unified rule-writing formalism and associated computational mechanism for a number of disparate kinds of process. Dictionary look-up, parsing, and morphological generation are all performed by the same rewrite-rule production system strategy. The mechanism has to be powerful enough to accommodate the most complex of those tasks, namely parsing. But to use all the power of a general rewrite mechanism for the relatively trivial tasks of dictionary look-up (involving only adding information to single arcs) and morphological generation (involving only string handling) may be seen as excessive. It does not matter so much for Météo because of the smallness of its scale, but these aspects of Météo's design are not necessarily desirable in larger systems.

12.5 Summary and discussion

Although Météo is clearly a second generation MT system, it departs from the classical architecture in interesting respects. Its linguistic strategies are well-motivated and finely tuned to the task on hand and from the computational point of view it is sophisticated, and for its time innovative.

In basic design, Météo is a direct system of a distinctive character. The limitation to the idiosyncratic nature of meteorological bulletins and the limitation to English and French translation in one direction led to a unique approach, which may well not be applicable in other circumstances. As we have seen, analysis lacks a morphological component, it is preceded by dictionary look-up which deals with any problems of 'idiomatic' or compound expressions, it is based on a mixture of English syntactic information and French semantic features, and the results are representations unique to the sublanguage of weather reports. Whereas analysis is bilingual, oriented to a specific target language, generation is wholly monolingual; but it is restricted to relatively minor syntactic manipulations and to morphological formations. It might, theoretically, be argued that Météo is in some senses an 'interlingua' system, with the intermediate representations between analysis and synthesis serving as 'pivot' texts. From this perspective, they are abstract representations of the information structure of bulletins which happen to have French labels. But the lexical units are not 'interlingual' concepts in any sense (e.g. *poudrerie* cannot be regarded as a unified (universal) concept). It is doubtful that these representations could be used as the basis for generating texts in any other language.

There seems to be no doubt that Météo is a clear example of a direct translation system, one which 'translates as it goes along'. The most typical 'direct' feature is the placing of lexical transfer before any syntactic analysis. Here it is done for perfectly good practical reasons. First, the vocabulary is highly restricted, there are very few homographs, and these are treated as category ambiguities (e.g *heavy* in section above). Second, the system is aimed exclusively at English–French translation: nothing is gained by postponing lexical transfer, and furthermore the knowledge of possible transfer ambiguities can be helpful in syntactic analysis. Third, source language lexical items are not actually replaced during analysis: the possible translation is added as extra lexical information to terminal items.

Météo is untypical of second generation systems in its weak modularity. There is no separation of the operations of analysis, transfer and generation; but it is modular in that the algorithmic and linguistic parts are separated: linguistic knowledge is expressed declaratively in an unordered set of rules, and the algorithmic knowledge which defines the procedural interpretation of the rules is captured by the Q-systems architecture. This means that extension of the sublanguage (e.g. to include gale warnings or reports on road conditions) could be achieved by adding linguistic rules and updating the dictionaries, without any change to the software. This was confirmed in practice by the rewriting and re-implementation of the software (on microcomputers and using the more efficient GramR language) without having to revise the linguistic foundations of the system.

What distinguishes Météo above all is the orientation to a particular sublanguage; this explains the preference for semantic analysis of structural relations and the use of features which would be unusual in general parsers ('blowing', 'falling'). The limitations of the approach are seen in part by the absence of any equally successful sublanguage MT systems and in the difficulties encountered by the TAUM group when they researched an English–French system for translating maintenance manuals for the hydraulic systems of jet aircraft (the Aviation project). There was a clearly identifiable sublanguage, but one which was much more complicated; in particular there were severe problems of multiple lexical, syntactic and pragmatic ambiguity in compound noun phrases and imperatives (19).

(19a) vertical stabilizer lower spar bar attachment fittings

(19b) Insert selector lever actuating arm in slot of selector lever.

The Aviation project differed in a number of respects from Météo (it was a transfer-based system and with much greater modularity), but the Q-systems approach proved ultimately unsatisfactory for this particular more complex sublanguage and the project came to an end in 1981.

The success of Météo is the best 'advertisement' for MT; it is successful because it is an exactly appropriate application: it performs satisfactorily and economically a translation task that humans do not enjoy doing, and which might otherwise not get done. The future of Météo, which is no longer being developed in any substantial way, is presumably to do its job until it is superseded by some better idea, or until the need to translate weather bulletins disappears. In fact, in recent years, some of the previous members of the TAUM group have been researching a system (RAREAS) which would generate bulletins in both languages directly from the raw meteorological data.

12.6 Sources and further reading

Basic descriptions of Météo are given by Chandioux (1976), Chevalier *et al.* (1978), summarised by Whitelock and Kilby (1983) and by Hutchins (1986: 228–231). Information about recent developments is given by Chandioux (1989a,b). For some details of the Aviation project, with further references, see Lehrberger and Bourbeau (1988); and for RAREAS see Kittredge *et al.* (1986).