CMPE 590 MACHINE TRANSLATION

RESEARCH PROJECT

09 MAY 2017

OGUZHAN MURAT CAKMAK

# METEO

# OVERVIEW

- INTRODUCTION
  - HISTORY
- SYSTEM DESIGN
  - INITIAL PROCESSES, THE TRANSLATION PROCESS, THE COMPUTATIONAL ASPECTS
- CONCLUSIONS
  - LIMITATIONS, ECONOMICAL ASPECTS, FUTURE, SUMMARY AND DISCUSSION
- REFERENCES

# INTRODUCTION

- TAUM(Traduction Automatique de l'Universite de Montreal) group in Mon-treal developed METEO which translates weather bulletins from English toFrench and has operated since 1977

# INTRODUCTION

- Q-SYSTEM
- DIRECT
- RULE-BASED
- NONINTERACTIVE
- PERFORMS ON SUB-LANGUAGE

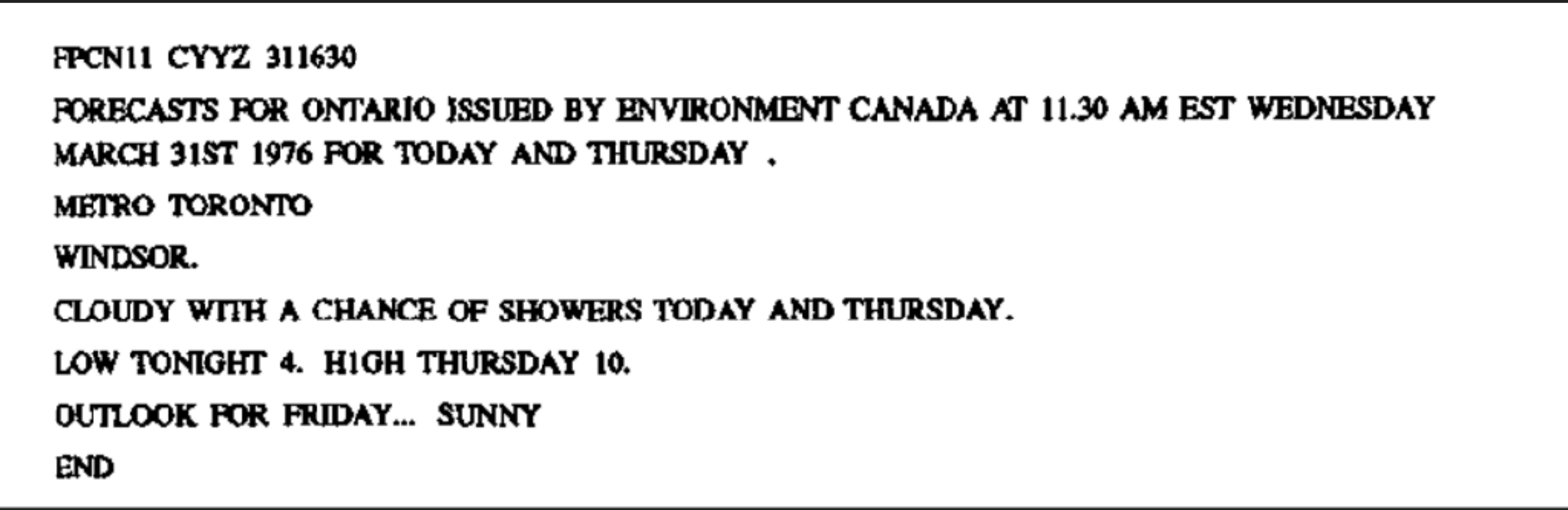
# INTRODUCTION/HISTORY

- 1965: CETADOL
- 1971: MT RISES
- 1975: CONTRACT FOR METEO
- 1976: FIRST PROTOTYPE
- 1977: ON LIVE
- 1984: METEO 2
- 1989: FRENCH-ENGLISH VERSION

# SYSTEM DESIGN

- Communication network sends the reports to the system.
- Pre-processes the data for METEO System.
- METEO detects any sentence it cannot translate and sends them to human translators.
- Reformats METEO output
- Transmit final version to the network.

# SYSTEM DESIGN



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. HIGH THURSDAY 10.

OUTLOOK FOR FRIDAY... SUNNY

END

## SYSTEM DESIGN/TRANSLATION PROCESS

- Three bilingual dictionaries for idioms, place names, and general vocabularies and three processing modules for the syntactic analysis of English, thesyntactic generation of French, and the morphological generation of Frenchexist in the system.
- Albeit its initial implementation, there was no need for a transfer module,because English reports have nearly the same structure as the French reports.Restricted morphological variation is the reason of the absence of the transfermodule. In the end, all lexical variants are in the dictionary.
- Meteorological reports have no pronominal reference, no relative clauses,no passive; and also all phrases are short. The semantic feature is applied tosolve problems due to the omission of prepositions and articles.

# SYSTEM DESIGN/TRANSLATION PROCESS/GENERAL OVERVIEW

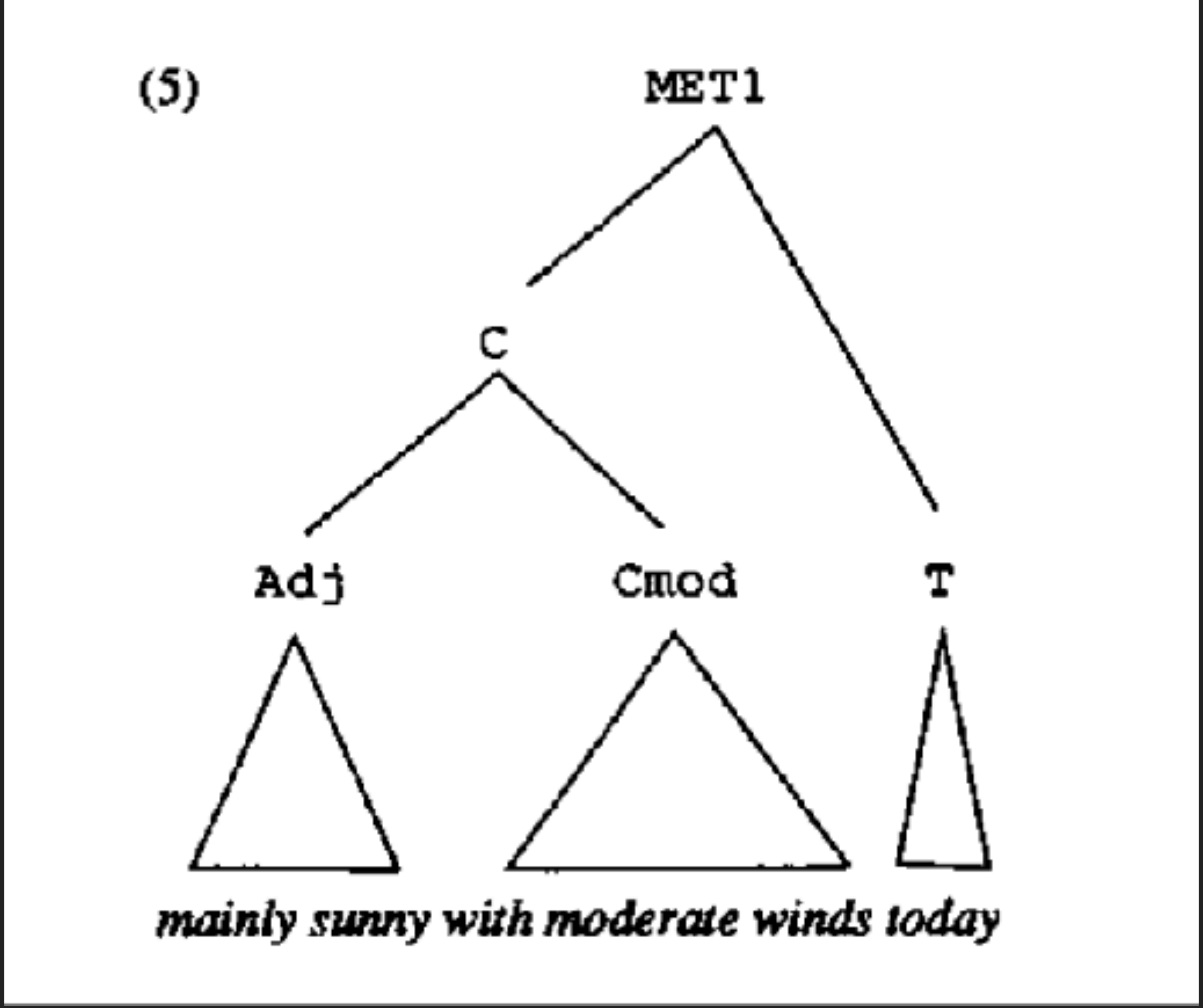
- The elimination of units which has errors due to some problems
- Translation of headings
- Expansion of abbreviations
- Recognition of idiomatic expressions
- Dealing with place names, dates, time, temperatures
- General dictionary look up
- Analysis of time references, locations, noun phrases
- Building of structures which represent a weather condition, the standard structures of units
- Elimination of partially analyzed units
- Syntactic generation
- Morphological generation

# SYSTEM DESIGN/TRANSLATION PROCESS/DICTIONARY LOOK UP

- (AMOUNT - N((F,MSR), QUANTITIE))
- N: Noun
- F: Feminine Gender
- MSR: Measure Noun

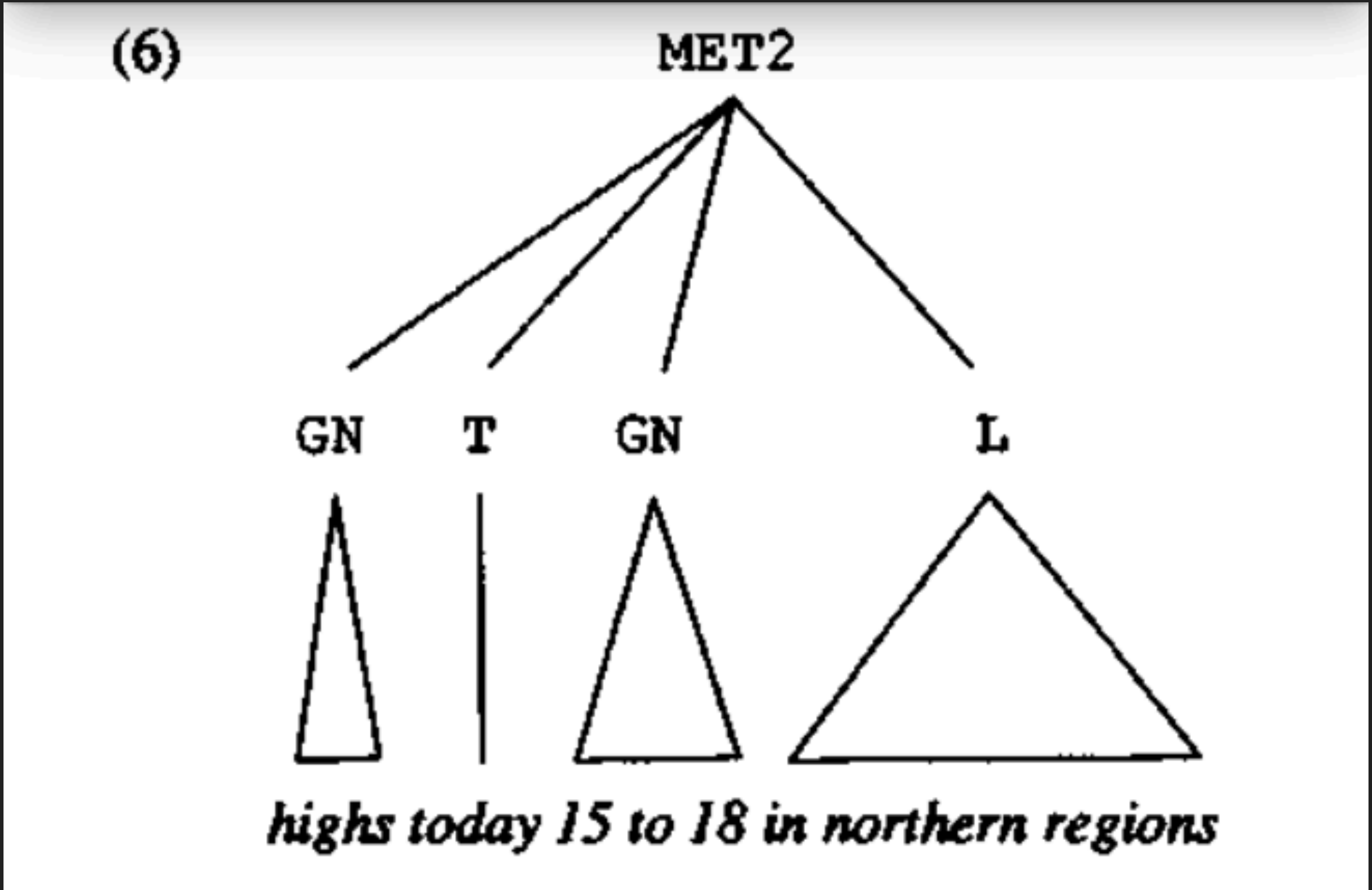
# SYSTEM DESIGN/TRANSLATION PROCESS/SYNTACTIC ANALYSIS

- MET1 (C({Adj, GN},[Cmod]),[T],[L])



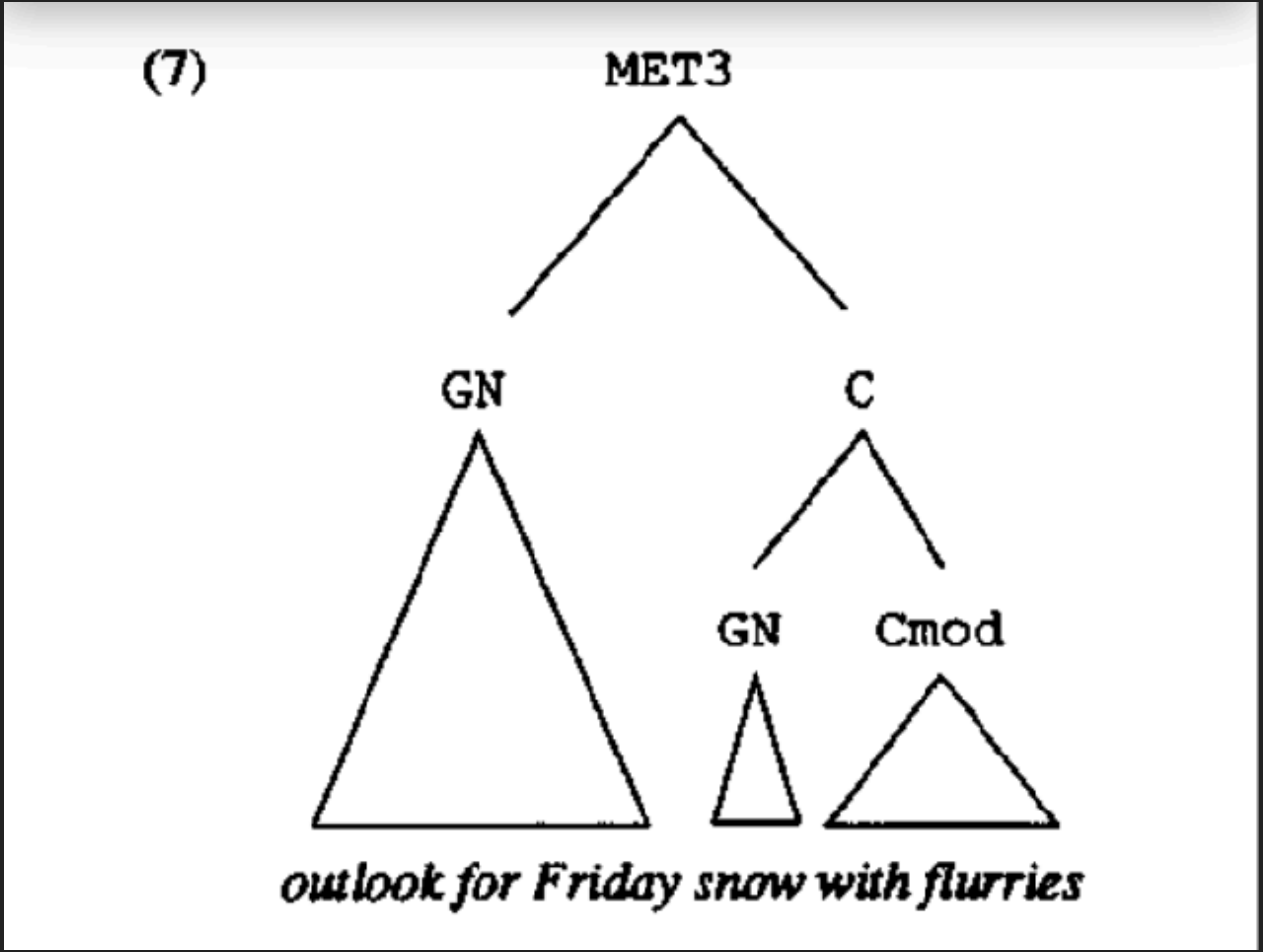
# SYSTEM DESIGN/TRANSLATION PROCESS/SYNTACTIC ANALYSIS

- MET2 (C({Adj, GN},[Cmod]),[T],[L])



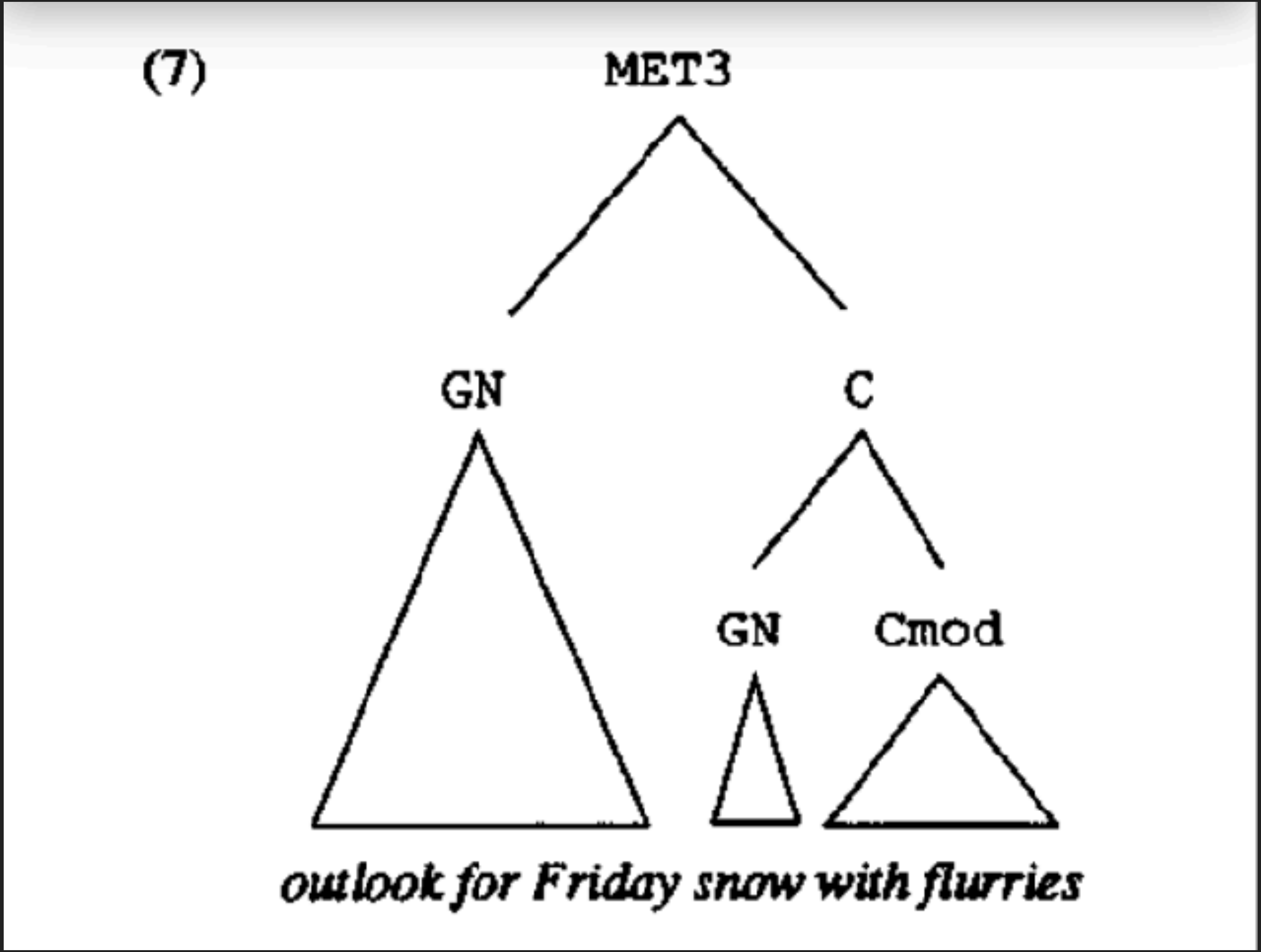
# SYSTEM DESIGN/TRANSLATION PROCESS/SYNTACTIC ANALYSIS

- MET3 (GN(outlook for T), C({Adj, GN},[Cmod]),[T],[L])



# SYSTEM DESIGN/TRANSLATION PROCESS/SYNTACTIC ANALYSIS

- MET3 (GN(outlook for T), C({Adj, GN},[Cmod]),[T],[L])



## SYSTEM DESIGN/TRANSLATION PROCESS/SYNTACTIC AND MORPHOLOGICAL GENERATION

- French word order is followed in the generation phase. Time and locationexpressions are after meteorological conditions. We locate adjectives afternouns.  Selection of correct preposition is considered because French hasdifferent articles for different genders.
- Ensuring the correct endings of adjective is also concerned at the final step.

## SYSTEM DESIGN/COMPUTATIONAL ASPECTS

- METEO is developed as a single monolithic system in the type of a production system. The chart format is the data structure. The arcs denote with feature bundle.  The order of application of grammar rule is not defined.The bottom-up parser is the production system architecture. The software is called Q-system and is regarded as the inspiration of language Prolog.

# SYSTEM DESIGN/COMPUTATIONAL ASPECTS/THE HARDWARE

- Control Data CDC 7600 is the mainframe computer of the Canadian Meteorological Center (C.M.C.). It carries out simulations of atmospheric condition. An automated process interrupts the simulations every five minutes and loads the METEO system, and it translates any messages since the previous interruption.

# SYSTEM DESIGN/COMPUTATIONAL ASPECTS/THE SOFTWARE

- A chart is a data structure of METEO, feature bundles are demonstrated on the arcs which start from these nodes. The early process of dictionary looks up developers primitive version of the chart by naming one arc for each reading of a lexical structure. If there is no ambiguity in a word, it has one arc at the end. If there is more than one meaning of a word, it ends up with the number of the arc which is the same as the number of meaning. For example, blowing snow has one arc, while ambiguous heavy has three arcs.
- The rule's 'pattern-match' part defines a sequence of the arc. Each rule actives "pattern-matcher" which look for a sequence of these arcs. When it finds, it builds a new arc which covers the whole sequence and labeled according to the rule's action part.

# SYSTEM DESIGN/COMPUTATIONAL ASPECTS/THE SOFTWARE

- It is crucial to show that the arcs used in the construction of a new arc. You can think of them like Transformers, each arc sacrifices them for something bigger. The new arc must shows which version of an ambiguous word is used during the creation. The important point is we should remove any arc because we can use those arc while applying other rules. Thanks to this approach, we keep our options open until the final complete analysis is done.
- Computationally there is no difference between the processes of analysis and generation. A generation has the same features and action as analysis. However, they are separate units. The unnecessary arcs are removed at the end of the process of analysis. The generation phase just deals with one arc. It is for computational efficiency, not for modularization of the translation procedure.

# SYSTEM DESIGN/COMPUTATIONAL ASPECTS/THE SOFTWARE/RULE FORMALISM

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

A tree 'WORSE's defined on the left-hand side and the right-hand side has a tree with two branches.

## SYSTEM DESIGN/COMPUTATIONAL ASPECTS/THE SOFTWARE/RULE APPLICATION

- The rule set is an unordered set of rules makes the Q-system procedure is a traditional production system. The state of the database determines which of the rules in the rule set may be applied next. We need to consider more than one rule. Theoretically, they run in parallel. Because of the computational restriction, we do it sequentially using a method called backtracking. We store slightly different copies of the databases and backtracks if the first rule comes ours as a wrong. This recursive nature looks obvious now, however, before language Prolog, it was a huge innovation.

# SYSTEM DESIGN/COMPUTATIONAL ASPECTS/THE SOFTWARE/RULE APPLICATION

- Unstructured approach to linguistic processing has many disadvantages, and METEO uses that relatively small rule set, and restricted text type to bypass them. The production system architecture looks a good fit for METEO; it may not be a good solution for general machine translation.

- The use of a single unified rule-writing formalism and the computation of different kinds of a process has possible drawbacks. The system is powerful enough to handle the most computational heavy task named parsing. This power is not utilized for the tasks like dictionary look-up and morphological generation. In METEO case, it is not a problem while in a large scale system it would be a huge problem.

# CONCLUSIONS

# CONCLUSIONS/LIMITATIONS OF THE SYSTEM

- METEO can not translate all the weather reports. The messages which the METEO system can translate are instance of regional forecast, maritime forecasts and forecast for farmers and boaters. Other information such as general synopses, and ice warnings are not intended to be translated by METEO.

## CONCLUSIONS/LIMITATIONS OF THE SYSTEM

- METEO can not translate all the weather reports. The messages which the METEO system can translate are instance of regional forecast, maritime forecasts and forecast for farmers and boaters. Other information such as general synopses, and ice warnings are not intended to be translated by METEO.

# CONCLUSIONS/THE SYSTEM'S FUTURE

- METEO became the face of the Machine translation because of its success. It creates useful outputs economically, and take away the mundane task that human does not would like to be part of.A kind of a ceiling is reached in terms of the linguistic model. Economically, it is not appropriate to invest a lot of extra resources to deal with rarer cases or errors in the original text. The next step in the era would be another technology. The future of METEO is not promising, it will be replaced by a superior machine translation system.

# CONCLUSIONS/SUMMARY AND DISCUSSION

- As a second generation, machine translation system METEO differs from the traditional architecture by finely tuned to the task of weather report translation and innovative computational point of view.

# CONCLUSIONS/SUMMARY AND DISCUSSION

- METEO works on the limit of the distinctive nature of meteorological bulletins and the limitation of the one direction translation from English to French. Since it is so specific, this approach does not scale up in other machine translation problems. There is no morphological component, the analysis is handled by dictionary look up which sees any problems 'idiomatic'. It depends on the mixture of English syntactic information and French semantic features. The analysis is bilingual, while generation is monolingual. There are minor syntactic manipulations and morphological translations. Since METEO has an intermediate representation between analysis and synthesis, it can be seen as an interlingua system theoretically. In this perspective, these representations are an abstract representation of the information structure of bulletins. However, it is the fact that lexical units are not interlingual' concepts. It is not sure that we can use these representations to translate into the other target language.

# CONCLUSIONS/SUMMARY AND DISCUSSION

- METEO is a real instance of a direct translation in any sense. The obvious reason is that it performs lexical transfer before any syntactic analysis. The domain's  limited nature makes it practical.
- Vocabulary is highly limited to the weather report domain.
- The system is exclusive to English-French translation: postponing gains nothing.
- The source language lexical items are not replaced immediately. It just accumulates extra information to the arcs.

# CONCLUSIONS/SUMMARY AND DISCUSSION

- METEO is an atypical second generation system since it has a weak modularity. Analysis, transfer, and generation modules are used in a monolithic manner. Algorithmic and linguistic parts are independent. Linguistic knowledge is structured as an unordered set of rules and declaratively. The algorithmic knowledge is defined by the Q-systems architecture software. Software remains the same if we would like to extend the sub-language. This assumption is proved while rewriting the software stack of the system.

# CONCLUSIONS/SUMMARY AND DISCUSSION

- Orientation to a particular sub-language is the most distinguishable aspect of METEO. The preferences is for semantic analysis of structural relations and the use of unusual features. METEO's unique domain differs it from the other domains such as aviation. TAUM group also initiates a project fro translation of maintenance manual for the hydraulic systems of jet aircraft. The sub-language is also more sophisticated than weather reports. Multiple lexical, syntactic and pragmatic ambiguity is found on the sub-language.

# CONCLUSIONS/SUMMARY AND DISCUSSION

- Orientation to a particular sub-language is the most distinguishable aspect of METEO. The preferences is for semantic analysis of structural relations and the use of unusual features. METEO's unique domain differs it from the other domains such as aviation. TAUM group also initiates a project fro translation of maintenance manual for the hydraulic systems of jet aircraft. The sub-language is also more sophisticated than weather reports. Multiple lexical, syntactic and pragmatic ambiguity is found on the sub-language.

# REFERENCES

- Hutchins, William John, and Harold L. Somers. An introduction to machine translation. Vol. 362. London: Academic Press, 1992.
- Lehrberger, John. "Automatic translation and the concept of sublanguage." Sublanguage: Studies of Language in Restricted Semantic Domains, R. Kittredge and J. Lehrberger, Eds. Berlin & New York: Walter de Gruyter (1982): 81-106.
- Thouin, Benoît. "The METEO system." Practical experience of machine translation (1982): 39-44.
- Mitkov, Ruslan, Somer, Harold. The Oxford handbook of computational linguistics. Oxford University Press, 2005.
- Langlais, Philippe, et al. "The long-term forecast for weather bulletin translation." Machine Translation 19.1 (2005): 83-112.

# METEO