

# **Support Tools are more Promising**

—

**Pierre Isabelle**

**Centre d'Innovation en Technologies de l'Information (CITI)  
Canadian Department of Industry**

**March 1994**

## **MT at the Canadian Translation Bureau**

### **From an Early Success...**

- 1977: implementation of TAUM-METEO
- Specifically designed for translating weather bulletins
- Improved over the years; but it was a success from the start!
- Current production: 20 million words per year
- Current production cost: less than 4 cents/word

## **...To a Series of Failures**

- 1981: TAUM-AVIATION project abandoned; deemed non cost-effective
- 1982: after a 6-month trial with system X it is concluded that "even in the most favorable cases, productivity with the system never exceeded 75% of the productivity of unaided translators"
- 1985: after a 4-month trial with system Y it is reported that "human translation proved to be between 30% and 34% faster than computer-assisted translation"
- 1992: after **5-year trial**, system Z is rejected on the grounds that production costs (around 70¢/word) are more than twice as high as with unaided translation

## **What went wrong?**

## **Weather Bulletins are Highly Atypical**

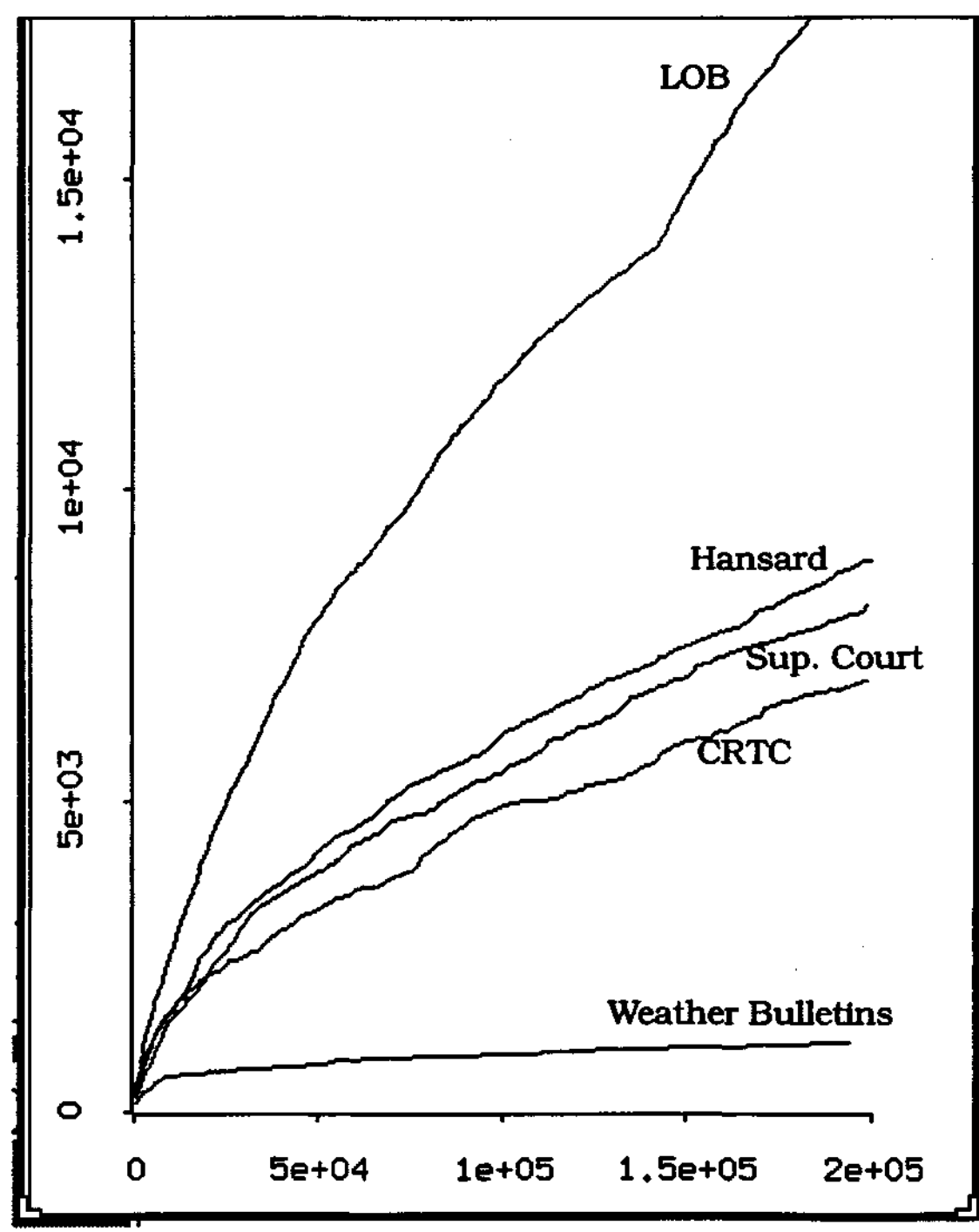
⇒ **Extreme linguistic simplicity:**

- less than 2000 lexical items, including 1000 proper names
- limited and closed inventory of semantic patterns (micro universe)
- stereotyped style

⇒ **Extremely high volume**

- **No other comparable situation has yet been identified in Canada**
- **No other comparable system implemented in other countries**
- **The bulk of the demand is for MUCH more difficult texts**
- **This is why clear-cut, uncontroversial success is so rare**

## Lexical Complexity of Sublanguages: Dictionary Size against Text Length



## **The State of the Art**

- Good-quality MT is only feasible in highly exceptional cases
- Poor-quality MT is found to be useless by most translation services
- This is why the current market share of MT is so small (no data, but my guess is *2 or 3 percent*)
- I don't see any evidence that this situation is about to change
  - ⇒ MT is best viewed as an excellent problem area for long-term research
  - ⇒ near-term goals should rather concentrate on less ambitious translation support tools

## **Translation Analysis**

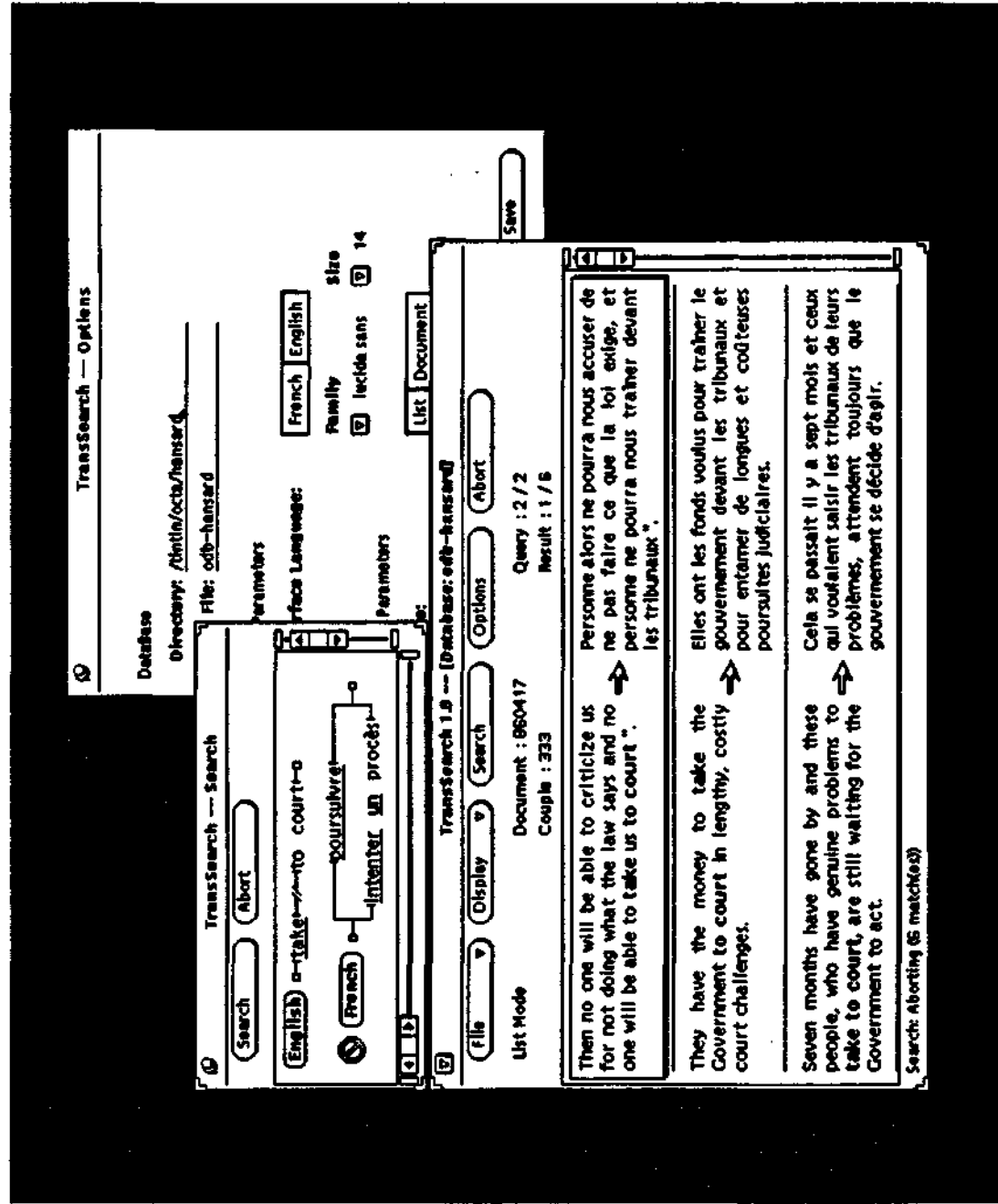
- MT focuses on the *production* problem; translation support rather calls for a focus on the *recognition* problem
- Translation analysis: reconstruction of the complex correspondences linking a source text and its translation
- In production mode strong translation models are mandatory, but in analysis mode, weak models can produce useful *partial* results (ex: sentence alignment)
- Corpus-based methods appear to be the natural choice
- Wide range of (new) applications in translation support:
  - translation memory
  - translation checking
  - translation dictation

## **Translation Memory**

- A translation analyzer is used to structure pre-existing translations in such a way that they become reusable in the production of new translation
- *Bilingual concordancing*: a simple but highly effective and universally useful kind of translation memory
- Allows users to search pre-existing translations for ready-made solutions to their translation problems
- *TransSearch*: the CITI's new bilingual concordancing tool



# TransSearch: a Screenshot



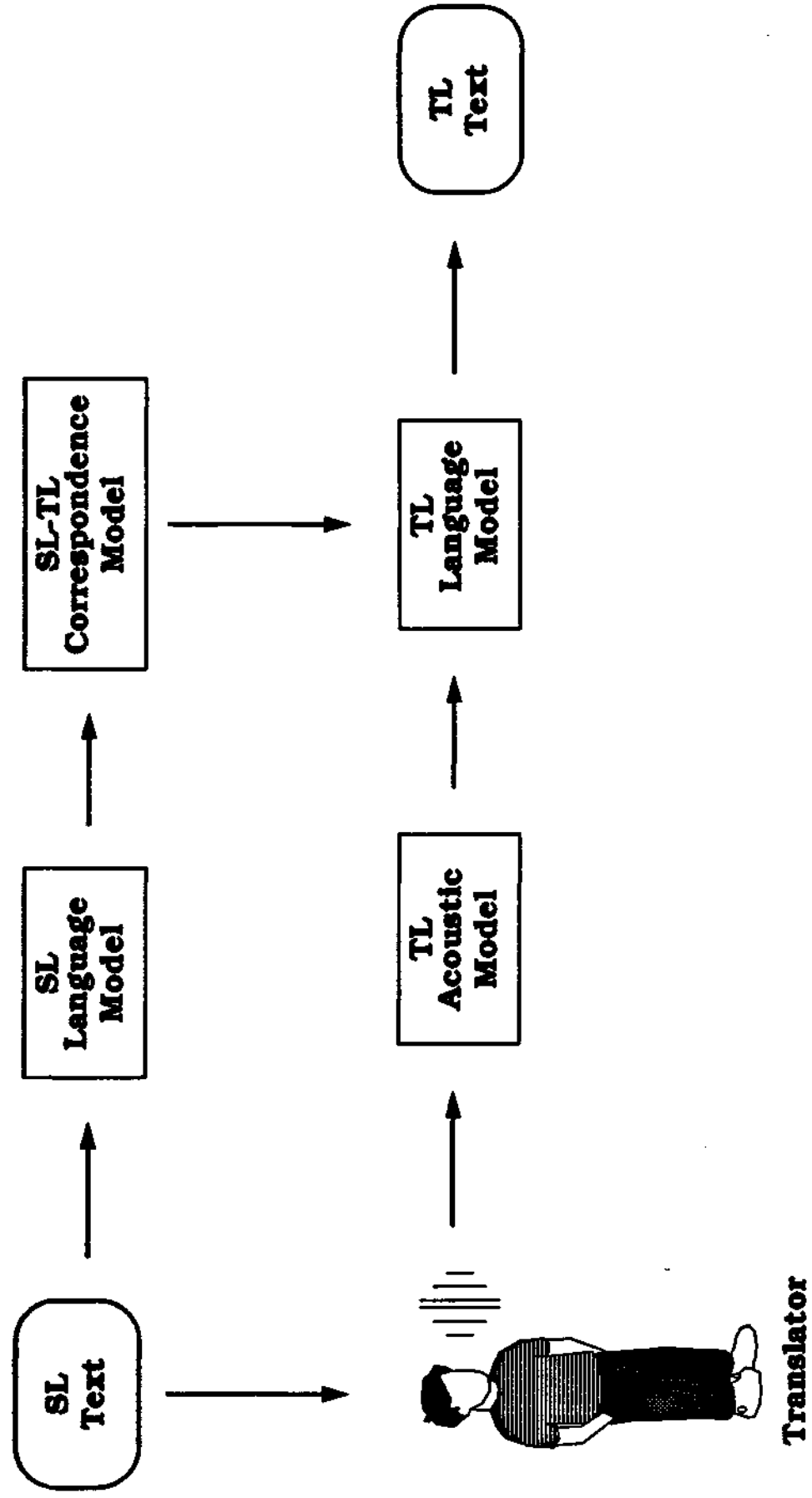
## Translation Errors

- Grammar checkers: verify that TT is correct wrt to TL; do not verify that TT has the right relationship with ST
- Correspondence errors such as *mistranslations* can only be detected through translation analysis
- Globally, translation quality is a thorny problem...
- But we can identify *simple properties* that most translators will want their translations to possess
  - ⇒ T is exhaustive; T does not contain deceptive cognates; etc.
- TA makes it possible to verify such properties
- *TransCheck*: our proof-of-concept prototype

## **Translation Dictation**

- The techniques of translation analysis can be used to develop high-performance translation-oriented speech recognition systems
- Combine two convergent sources of information:
  - speech signal
  - predictions made by a (weak) translation model applied to written the source text
- Analyze a translation where the target is spoken rather than written
- TransTalk: a prototype translation dictation system being jointly developed by CITI and CRIM

# The TransTalk Model of Translation Dictation



## Conclusions

- Most translation services consider that MT cannot help them at this time, and that MT is best viewed as a problem area for long-time research
  - However, in our experience, most translations services are extremely interested in new kinds of support tools
  - Corpus-based methods lend themselves very well to the production of (partial) analyses of human translations
  - Such analyses open up the way for increasingly powerful translation support tools, including
    - translation memories
    - translation checking
    - translation dictation
- ⇒ The future looks bright for translation support technology