



Automated Translation System

Job Bank

August 2006



Table of Contents

- Introduction
- Creating a Job Offer
- Automated Translation System Components
- Translation Quality Evaluation
- Revision Rate/Results
- Next Steps

Job Bank Site

(www.jobbank.gc.ca)

- Formed in mid '60's and has evolved to automated system
- Employers can post jobs 24/7
- >270,000 employer accounts
- >60 million user sessions recorded in 2004-05.
- One million new offers posted in 2005-06. Growth: >30% per year
- Job offers received by fax, e-mail, Internet and telephone
- **Must** be published by next EOB
- Maximum time for revision is 4 hours

Legislative and Political Background

- Section 60(1) of the Employment Insurance Act : “... a national employment service ... employment opportunities ... suitable employment ... suitable workers.”
- Official Languages Act (OLA), obligates translation of all job offers
- In 2003, approval of a 3-year translation action plan leading to an automated translation system.
- In a 2005-2006 report, the machine translations of some job offers are still of questionable quality.
- Following negative media coverage, Job Bank was called in June 2006 to the House of Commons Standing Committee on Official Languages to discuss the translation of job offers on Job Bank. Result: hire more ‘post-editors’ or revisors

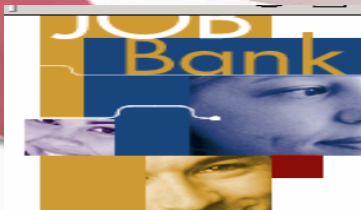
Building the Translation System

- In 2002 - rudimentary automated system: very poor results.
- Estimated cost for human translation of all job offers (450,000 jobs at the time) without relying on technology was *\$25 million*.
- Requirements and solutions studies conducted.
- Recommendations: Job Bank developed pre-translated checklists, an automated system to translate free text and hired a team of revisors.

Procurement Process

- 2002-2003
 - Identification of translation requirements
 - RFP for integrated solution: no successful bids
- 2003-2004
 - Second RFP with separate components
 - Success: for MT and TM
- 2004-2005
 - Integration of memory; continued update to MT dictionaries
- 2005-2006
 - RFI for grammar and spell checker; no successful bids
 - Internal research for available products; procurement
- Present
 - Final integration of grammar and spelling checker
 - Continued improvements: changes to system to implement the new approach to the translation of job offers

Creating a Job Offer



Pre-translated
checklists -
examples

Yuki & Wolf Writing and Translating (#304488)
Advertising writer NOC #5121

Create a Job Advertisement

Mandatory fields are indicated by a ★.

Step 4 of 9: SPECIFY ANTICIPATED STARTING DATE

★ Select one of the following:

- As soon as possible
- Specified date:

Step 5 of 9: SPECIFY SALARY INFORMATION

★ Select one of the following:

- Salary :**
 for hours per week.
- Depending on Experience :**
Salary from to for hours per week.
- To be Negotiated**
(Employer undertakes to ensure compliance with all employment standards statutes)
- Remuneration Based on Commission Only**
(Employer undertakes to ensure compliance with all employment standards statutes)

• Check all that apply:

| | |
|---|-------------------------------------|
| <input type="checkbox"/> Bonus | <input type="checkbox"/> Piece Work |
| <input type="checkbox"/> Gratuities | <input type="checkbox"/> Commission |
| <input type="checkbox"/> Other Benefits | |

Step 6 of 9: SPECIFY TERMS OF EMPLOYMENT

★ Select one of the following:

Permanent Temporary Casual Seasonal

★ Select one of the following:

Full Time Part Time Part Time leading Full Time

• Check all that apply:

| | | |
|---|--------------------------------|-----------------------------------|
| <input type="checkbox"/> On Call | <input type="checkbox"/> Shift | <input type="checkbox"/> Overtime |
| <input type="checkbox"/> Weekend | <input type="checkbox"/> Day | <input type="checkbox"/> Night |
| <input checked="" type="checkbox"/> Evening | | |

Mandatory fields are indicated by a ★.

Creating a Job Offer



Free text

Yuki & Wolf Writing and Translating (#304488)

Advertising writer NOC #5121

Job Advertisement

SPECIFY OTHER INFORMATION

If needed, this free-text space is provided for you to include further details about your job ad. This text is fed through a machine translation system which will produce a literal translation of your wording. All errors and ambiguities contained in the original text will then be reflected directly in the translated language. Please follow the helpful hints (below/above) to produce clear and concise sentences.

1. Write short, complete sentences.
2. Avoid abbreviations and acronyms.
3. Use capital letters only for proper names, official titles and all nouns that should be capitalized (language, place names, days of the week, etc.)
4. Avoid typographical and grammatical errors. Use correct punctuation and verify that all spaces have been included between words, characters or symbols.
5. Please make use of Job Bank's spell-check feature.

• Other Information:

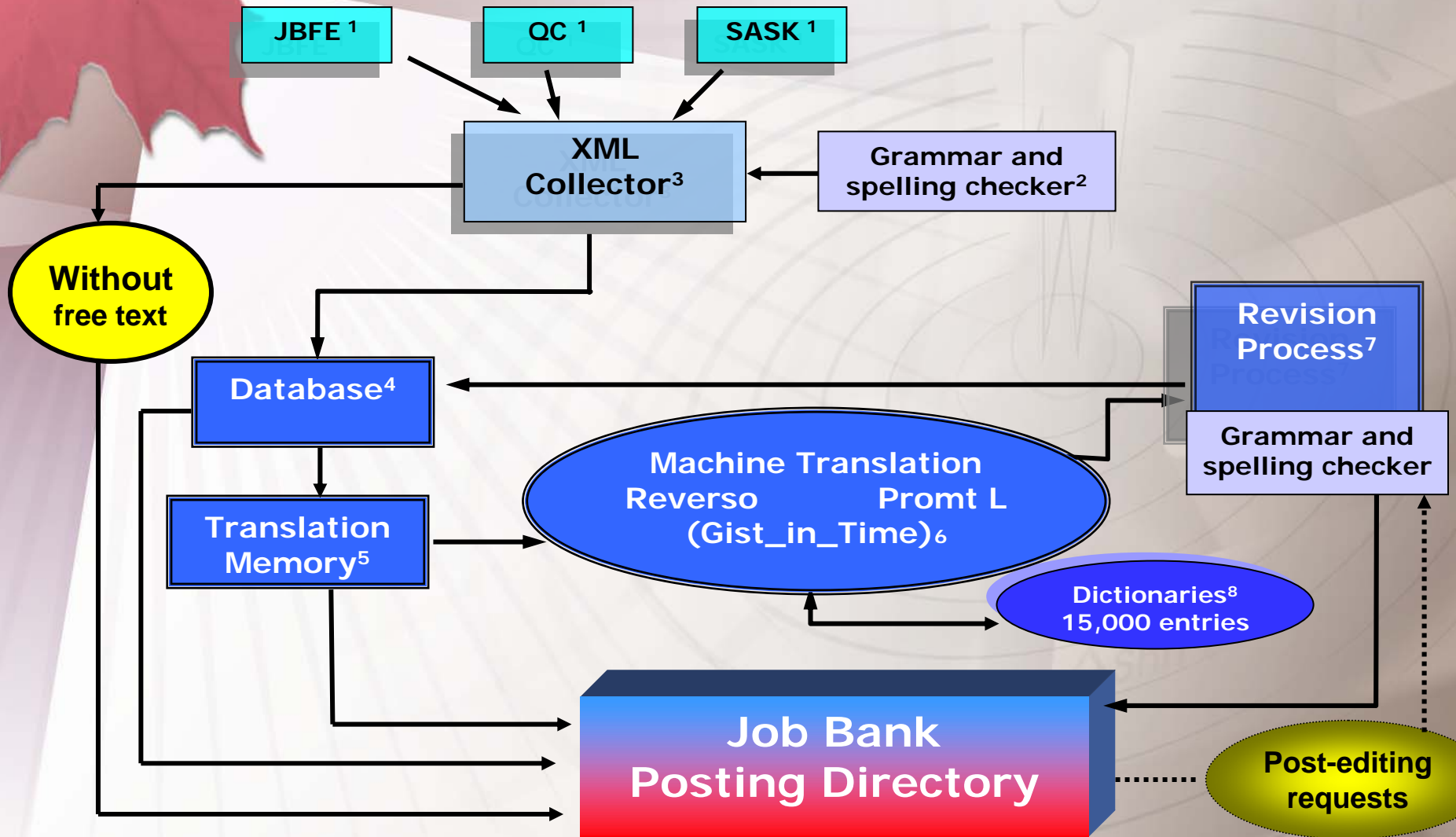
(max.
300
char)

0

CANCEL

CONTINUE

Translation System Architecture



Specialized Dictionaries

Dictionaries⁸
15,000 entries

- Development of a multiple dictionary structure - 10 specialized dictionaries (employment-related words and expressions) activated simultaneously
- Extraction of mistranslated terms from the National Occupational Classification (NOC) as well as from offers revised by the revision team
- Addition of idiomatic expressions that cannot be translated literally
- Creation of a dictionary of geographical names (1200)
- Dictionary updates each quarter

Base dictionary

Basic words: one word, one equivalent
(sometimes two)

Examples :

- performance : exécution (performance)
- filet : net
- mignon : good-looking

Literal vs. Correct Translation

Examples :

- Commerce de gros et de détail : *business of big person (main part) and detail* - *wholesale and retail trade*
- Manger un filet mignon : *eat a good-looking net* - *eat a filet mignon*
- Go to Swift Current : *allez à Courant Rapide* - *allez à Swift Current*
- The musician will give a performance : *le musicien donnera une exécution (performance)* – *le musicien donnera un spectacle*

‘Equivalents’ Dictionary

- Contextual phrases ensure more appropriate translation.
 - E.g. international trade : commerce international
 - construction trade : métier de la construction
 - trade certification : certificat d'aptitudes professionnelles
 - trade contractor : entrepreneur spécialisé
 - trade periodical : périodique professionnel
 - trade show : salon professionnel
 - trade economist : économiste du commerce
 - trade union : syndicat
 - perform a duty : exercer une fonction
 - perform a task : exécuter une tâche

Database

Database⁴
300,000 offers

- Archiving offers and their revised equivalents in a database
- Automatic posting of offers that are identical along with their translation
- 38% of offers posted are reproduced by the database

Translation Memory

Translation
Memory⁵
1,450,000 words

- Creation of a corpus of full sentences extracted from job offers
- Automatic reproduction of translations of repetitive sentences in job offers
- The corpus currently contains 1,450,000 words (source and target text)
- Approximately 16% of offers posted are reproduced by the translation memory
- Update of new corpora every quarter

Text from the database

Translations

Education/Éducation

Formation reconnue

Recognized training

Edit

Competences description/Description des compétences

Les personnes recherchées devront composer facilement avec le travail d'équipe. Avoir un sens éveillé du service à la clientèle et posséder des aptitudes à la vente-conseil. Reconnaissance de l'expérience de travail et des heures de formation.

The ideal candidates will have to work easily as a team. Have strong customer service, sales and advice skills. Will recognize work experience and training hours.

Edit

Additional Information/Précisions additionnelles

aucun appel téléphonique

No phone calls

Edit

Duty Description/Description des tâches

Les personnes recherchées devront prodiguées les soins du visage, de pieds, manucure et il serait un atout de posséder une connaissance des soins corps tel que enveloppements et exfoliations corporelles.

The ideal candidates will perform treatments on the face, feet, and manicures. Knowledge of body care, body wraps and physical exfoliations is an asset.

Edit

Other Information/Informations complémentaires

Dû à la croissance de notre entreprise, nous sommes présentement à la recherche de plusieurs candidats pour différentes régions du Québec, dont Montréal, Laval, Tremblant, Bécancour, St-Foy, La Malbaie et Gatineau.

Due to our company's growth, we are currently seeking several candidates in various regions of Quebec, including Montreal, Laval, Tremblant, Bécancour, St-Foy, La Malbaie and Gatineau.

Edit

Text from the memory

Duty Description/Description des tâches

CIM (Consumer Impact marketing) est le chef de file dans l'impartition des forces de ventes et d'exécution marketing. Reconnue comme étant l'une des 50 sociétés les mieux gérées au Canada, CIM est le partenaire stratégique de plusieurs entreprises reconnues dans le monde pour l'impartition des forces de vente et de merchandisage, la gestion d'événements, de promotions et des ventes directes. Afin d'aider un de nos clients à promouvoir ces services résidentiels télévisuel, internet et bientôt téléphonique, nous recherchons 1 représentant(e) des ventes pour couvrir le territoire de la région de Magog et environs. Ce représentant(e) aura comme responsabilité de visiter les domiciles de leur territoire pour offrir les différents services.

CIM (Consumer Impact Marketing) is the leader in the third party outsourcing of sales force and marketing execution. Recognized as being one of the 50 best managed companies in Canada, CIM is the strategic partner of several companies recognized worldwide, outsourcing sales and marketing forces, events management, promotions and direct sales.

To help one of our customers to promote these services television, internet and soon phone résidentiels, we are looking for 1 representative of sales to cover the territory of the region of Magog and neighborhood. This representative will have for responsibility for visiting the places of residence of their territory to offer the various services

Memory



Dictionaries



Text from the dictionaries

Competences description/Description des compétences

Dynamique et motivé Capacité de soulever des boîtes d'un maximum de 25 livres Permis de conduite en règle et accès à un véhicule fiable Petit espace d'entreposage au domicile pour la gestion du matériel et des outils promotionnels Accès à un ordinateur et à un courrier électronique Capacité à travailler en toute indépendance Ordinateur avec accès à Internet et voiture requis Prévoir espace de rangement pour le matériel

Dynamics and motivated Capacity to lift(raise) boxes of a maximum of 25 books(pounds) Licence of driving(behaviour) in rule(ruler) and access to a reliable vehicle Small space of storage in the place of residence for the management of the material and the promotional tools Access to a computer and to an e-mail Ability to work in any independence Computer with access to internet and car required Foresee storage space for the material

```
Dynamic and motivated;  
Able to lift boxes of a maximum of 25 lbs;  
Valid driver's license and access to a reliable  
vehicle;  
Small storage space at home for materials and  
promotional tools;  
Access to a computer and to e-mail;  
Ability to work completely independently;  
Computer with Internet access and vehicle required;  
Plan to have storage space for materials.
```

60 Word(s) 367 Characters

Revision Interface

Revision
Process⁷

- Prioritized revision of offers for bilingual communities and positions (our main target)
- Saves revised texts in a database
- P / T's can do their own revisions
- Integrates grammar and spell checker

Post-editing

Our post-editing team consists of:

- 5 full time and 12 part-time post-editors
- Human post-editing rate: 17.5% of total offers

Current data

➤ Evaluation of the quality of machine translated texts using specialized dictionaries

English-French:

- Base dictionaries: 56.9 %
- Specialized dictionaries: 69.0 %

Improvement: 12.1 %

French-English:

- Base dictionaries: 63.0 %
- Specialized dictionaries: 75.9 %

Improvement: 12.9 %

Evaluation of the contribution of the Translation Memory

Calculation of the ratio of sentences found in the memory, multiplied by the quality coefficient (95%)

Automated translation quality index =

(% of sentences found in the translation memory X 95%) +

(% sentences translated by the translation

machine with dictionaries X quality of machine translation)

➤ English-French:

$$(16\% \times 95\%) + (84\% \times 69\%) = 73\%$$

➤ French-English:

$$(16\% \times 95\%) + (84\% \times 75.9\%) = 79\%$$

Evaluation of Machine Translation

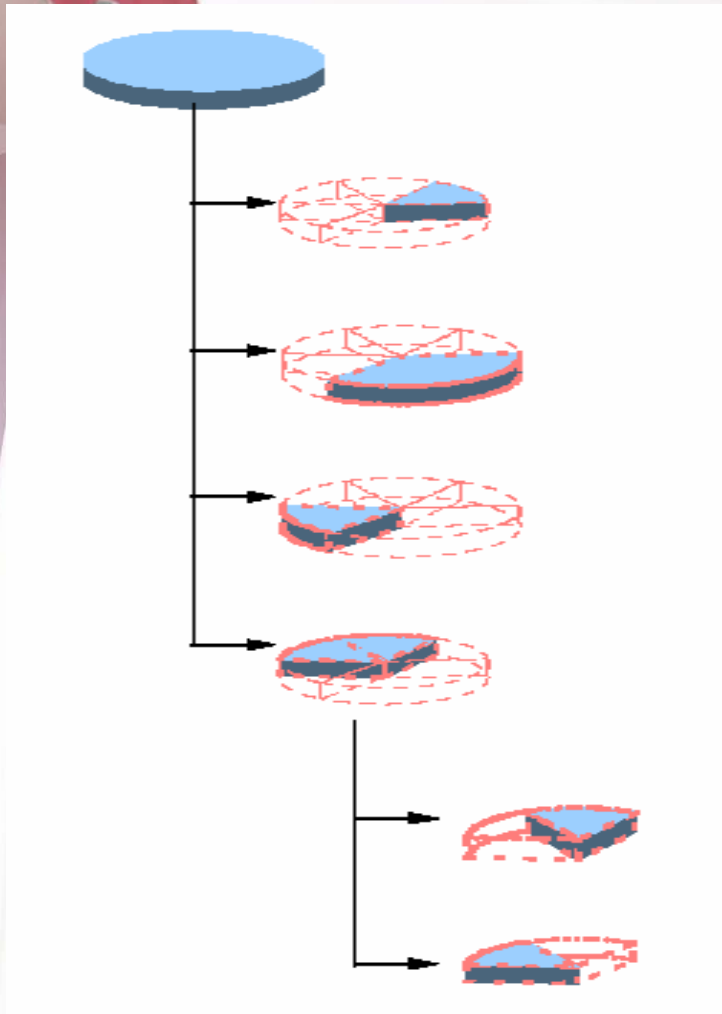
- Methodology for evaluating the quality of translated texts

- Evaluation grid :
 1. Intelligibility (text clarity)
 2. Adequacy (equivalence between texts)
 3. Language quality (spelling, grammar)
 4. Symbols and typography (“ ”, (), \$, ...)

Evaluation grid

| | INTELLIGIBILITY | ADEQUACY | LANGUAGE QUALITY | SYMBOLS AND TYPOGRAPHY |
|--------------------|--|--|--|---|
| DEFINITION | Clarity, comprehensibility and logic of the text. | Semantic equivalence between the source and target texts. | Respect for spelling, grammar (agreements, conjugations, etc.) and syntax (sentence structure) in the target language. | Respect for typographical rules in the target language: symbols, capitalization, punctuation, date and hour formats, parentheses, quotation marks, currency, etc. |
| METHODOLOGY | Read only the target sentence. Answer the following question: Is the text comprehensible at first reading? | Read each sentence, comparing the source and target. Determine the number of ideas in each sentence, count those which are well translated and assign points accordingly. | Read the target sentence. Count the number of spelling, grammar and syntax errors. | Read the target sentence. Count the number of errors. |
| SCORING | 0 = Incomprehensible text 1 = Certain elements are understandable, but the general meaning is lost 2 = The general meaning is understandable, even if some elements are unclear 3 = The text perfectly understandable | 0 = Nothing is properly translated 1 = A third of the ideas are properly translated 2 = Two-thirds of the ideas are properly translated 3 = All ideas are properly translated | 0 = 3 or more errors 1 = 2 errors 2 = 1 error 3 = No errors | 0 = 3 or more errors 1 = 2 errors 2 = 1 error 3 = No errors |
| WEIGHTING | 30% | 40% | 20% | 10% |

Offers posted with linguistic quality assured



(100%) **Total offers translated**

(15%) **Without free text**

(38%) **Recurring texts database**

(16%) **Translation memory**

(31%) **Machine translation with dictionaries**

(17.5%) Post-edited within 4-hour turnaround time

(13.5%) Not post-edited (turnaround time elapsed)

Next Steps

- Implement post-editing of *all* offers
- Complete grammar and spell checker integration
- Ongoing refinements to dictionaries, TM and database
- Cont'd improvements to the pre-translated NOC checklists
- Cont'd improvements to the translation system
- Perhaps apply technology to other official texts??



Questions?