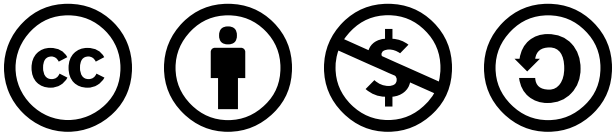
Translating for Canada, eh? (version 2)

Translating for Canada, eh? (version 2)

Lynne Bowker

University of Ottawa



Translating for Canada, eh? (version 2) Copyright © 2023 by Lynne Bowker is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, except where otherwise noted.

Contents

- Introduction
- Glossary
- Chapters
- Translation and Localization
- Term Banks
- Bilingual Concordancers
- Tools for Investigating Regional Variation in Languages
- Machine Translation Tools
- ChatGPT
- Language Portals
- Concluding Remarks
- Acknowledgements
- About the Author

Introduction

Welcome to the Open Educational Resource (OER) on translating for a Canadian audience. In the multilingual country of Canada, both English and French have equal status as official languages. In addition, there are more than 60 Indigenous languages in Canada, and in some regions, some of these languages have official status too. What's more, many heritage languages are used by immigrants who have arrived in Canada from all around the world. Through the Census Program, Statistics Canada gathers a variety of data about the range of languages used in this country, all of which point to Canada's rich history of multilingualism and translation.

Who is this resource for?

This resource is intended for newcomers to the field of translation in Canada. This includes students who are training to become translators, recent graduates who are just launching their careers in the language professions, students of other disciplines who are curious about translation, or even members of the general public who are interested in learning about this field. Of course, it's not absolutely essential for you to be a ***Canuck*** to learn from this resource, but the focus will be on exploring tools and resources that will be helpful when translating between English and French for a Canadian audience in particular.

It's common knowledge that there are differences between the varieties of English spoken in the United States and the United Kingdom. For example, in the US you take an ***elevator***, whereas in the United Kingdom you take a ***lift***. Likewise, the varieties of French spoken in France and Belgium have some differences too: in France, the number that comes after 89 is ***quatre-vingt-dix***, while in Belgium this number is ***nonante***. Well, Canada also has its own special flavour of both English and French, which means that if you are translating for an audience in this country, it pays to know your ***Timmies*** from your ***toonies***, otherwise you could find yourself in ***double double*** trouble!

What's in this resource?

This is not a comprehensive course book on how to do translation, nor is it an exhaustive description of every type of tool that may be useful for translation. Rather, it's resource that seeks to familiarize those who are translating for a Canadian audience with a selection of tools and resources that have a Canadian

connection. The primary focus is on tools and resources that support translation between the distinctly Canadian varieties of English and French, although the section on Language Portals contains a link to resources for some of Canada's Indigenous languages.

In addition, because this is an OER for students or those just starting out in the language professions, it only covers tools and resources that are **freely available online**. No information is provided about commercial products that are offered for purchase or by subscription.

Finally, this OER is not intended to be a detailed user guide for the tools and resources that it presents. Rather, the OER is meant to function as a sort of "quick reference" to raise awareness about tools and resources that may be useful when translating for a Canadian audience. Therefore, it introduces some highlights and basic features, and it provides a series of short practical exercises to get users started, but users are then referred to the user guides provided by the developers to obtain more details.

The **French version of this ebook** is available here!

2

Glossary

Bilingual concordancer: a tool that allows a user to enter a search term and to retrieve all examples of that search term from a parallel corpus. The source language sentences (containing the search term) are displayed in one column and the target language translations of those sentences are displayed in a parallel column.

Corpus (plural = corpora): a large collection of texts in electronic form that can be processed with the help of tools such as concordancers.

Interpretation: the transfer of a message from one spoken or signed language to another.

Locale: a combination of a language and region and their linguistic and cultural, conventions and preferences (e.g. French Canada, French Belgium).

Localization: Translation that goes beyond a given language (e.g. English, French) and takes into account a regional language variety and its cultural conventions and preferences (e.g. Canadian English, Canadian French).

Large Language Model (LLM): a type of artificial neural network designed for language processing tasks (e.g., question-answering, text generation, summarization, text simplification, and translation). These models contain a huge corpus of texts that a tool such as ChatGPT can consult to identify linguistic patterns and use as examples.

Machine learning: a branch of artificial intelligence which uses data and algorithms to imitate the way

that people learn and which gradually improves its accuracy.

Machine translation (MT): a type of software that takes a source text written in one language and produces a target text in another language.

Neural machine translation (NMT): a corpus-based approach to machine translation that uses artificial neural networks and machine learning techniques. The software is provided with an extremely large parallel corpus of texts and their translations and it consults these previously translated texts and "learns" to translate new texts.

Parallel corpus (plural = parallel corpora): a bilingual corpus that contains source texts aligned with their translations. The texts are broken down into sentences and each sentence in the source text is linked to its translation in the target text. A bilingual concordancer can be used to retrieve the source language sentences containing a search term and to display these alongside the equivalent target language sentences.

Portal: a specially designed web-based platform that collects information from different sources and presents it in a uniform way.

Prompt: the input or query provided by a user to initiate conversation or request assistance from an AI tool such as ChatGPT. The prompt typically consists of a question or instruction.

Prompt engineering: the process of creating input (e.g., a question or an instruction) for a tool based on large language model (LLM) (e.g., ChatGPT). The more specific and clear the prompt, the more likely you are to get useful results.

Rule-based machine translation (RBMT): the earliest approach to machine translation that attempted to get computers to process language using large bilingual dictionaries and complex sets of grammar rules.

Source language: the language of the original message that needs to be translated (i.e., the start language).

Source text: the text containing the message that needs to be translated (i.e., the start text).

Target language: the language into which the translation takes place (i.e., the end language).

Target text: the translated text (i.e., the end text).

Term: a word or expression from a specialized field of knowledge.

Term bank: a dictionary-like database that contains terms along with information about the terms, such as definitions, synonyms, equivalents in other languages, and notes about how terms are used (e.g. whether they belong to a specific language variety, such as Canadian English or Canadian French).

Translation: the transfer of a written message from one language to another.

Translation and Localization

Translation

So what is translation, exactly? Understood broadly, translation is the act of taking a message that has been created in one language, such as English, and expressing it in another language, such as French. When used in this broad sense, translation can encompass written, spoken and signed modes of language, and sometimes even a mixture of these. However, translation is usually understood more narrowly to refer to the transfer of a *written* message from one language to another, while interpretation is more commonly used to refer to cases where the message is spoken or signed. In this resource, we'll be using the term translation in the narrower sense of working with written texts.

Localization

For some types of texts, it can be sufficient to simply translate between languages, such as English and French. Most languages do indeed have a sort of standard variety that can be understood by the vast majority of speakers of that language. However, sometimes it's important to go beyond this high level and to focus more specifically on a particular language variety, such as Canadian English or Canadian French. When a text is customized to take into account a particular language variety and regional culture, the process is often referred to as *localization* rather than simply translation. When localizing a text, a translator needs to take into account not only the language but also the region, and the language and region together make up the *locale*. So English is a language, but Canadian English, Irish English, Australian English, US English, UK English, etc. are locales.

Try it! Can you spot the Canadian expressions?

Try these fun quizzes to test your knowledge of Canadian expressions:

- British and Canadian English: The sentences contain British words or expressions within quotation marks. Choose the corresponding Canadian word or expression.
- Expressions françaises et canadiennes: À vous de remplacer les expressions françaises placées entre guillemets par des expressions plus couramment utilisées au Canada.

Source: Language Portal of Canada, Translation Bureau, Public Services and Procurement Canada.

If you take a look at the website for a multinational company or an organization with a strong international presence, you will likely find that they have not just translated their website into different languages (e.g. English, French, Spanish) but they have gone a step further and produced localized versions aimed at different locales. The Coca-Cola Company is an example of a company whose website has been localized for a variety of locales.

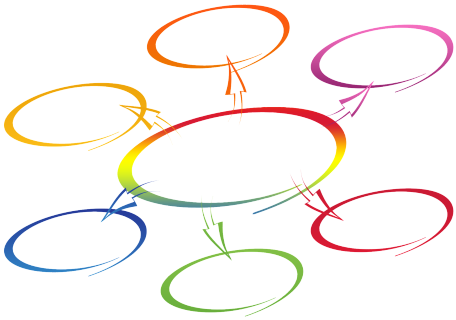
Try it! What differences do you see in the localized websites?

1. Go to the website of The Coca-Cola Company.
2. Compare two or more versions of the website that have been localized into different varieties of the same language (e.g. the versions for Australia, Ireland and Canada).
    - What do the different localized versions have in common?
    - What differences can you observe?
3. Now compare different versions that have been localized into different languages within a single country or region (e.g. the versions for English Canada and French Canada, or for Dutch Belgium and French Belgium).
    - What do you notice here in terms of similarities or differences?
4. Reflect on how two regions can share an overall language but still have some local linguistic differences, whereas other regions can share elements of a culture but still use different languages. Studying the different versions of these websites helps to emphasize that localization goes beyond language to incorporate regional linguistic and cultural content.

It's not difficult to appreciate that, in many circumstances, a reader would prefer to engage with a text in the language variety that they know best. Therefore, from a business perspective, it often makes sense for companies (or governments or other organizations) to localize their texts rather than simply translating them. But in order to produce a localized version, it is very helpful for translators to be able to consult resources or use tools that can help them to identify and apply the terms and expressions that belong to the language variety in question. So what tools and resources are available to help translators produce texts for Canadian audiences? Let's take a look at some language resources and tools that are **Made in Canada**, eh?

2

Term Banks



In the case of English and French, it is mainly the vocabulary that distinguishes the Canadian varieties of these languages from varieties spoken in other parts of the world. The grammar or syntax does not change very much, but there are definitely some words or expressions that are typical in Canada and that are used less frequently elsewhere. Therefore, when translating for a Canadian audience, it's a good idea to use the vocabulary that is common to the region. But where can a translator go to find such terms?

One useful type of resource is known as a term bank. Term banks tend to focus on language that is a bit more specialized than the type of everyday language that is found in a dictionary. While dictionaries hold a record of everyday language, term banks contain information about the concepts that belong to specialized fields of knowledge, such as law, engineering, education, sports, arts and culture, agriculture, tourism, economics, health, and… well… pretty much any area of specialized knowledge. And this actually makes sense because most of the texts that need to be translated are also specialized to some degree, so term banks are a very handy source of information for translators.

In addition to the terms themselves, term banks may also contain definitions, examples of the term in context, an indication of any synonyms or equivalents in other languages, as well as other observations, such as usage information (e.g. whether a term is preferred in a certain geographic region or organization). With regard to usage information, a term bank may actually make recommendations about which terms to use and which terms to avoid. Being able to determine that a certain term is preferred for a Canadian audience or for an organization such as the Canadian government (which is a major employer of translators in Canada!) is valuable.

TERMIUM Plus

One of the biggest term banks in the world, *TERMIUM Plus* is maintained by the Government of Canada's Translation Bureau. It contains millions of term records in English and French (with some records in Spanish and Portuguese). In addition to the terms themselves, these records may contain information such as definitions, synonyms, abbreviations, and usage examples. The records may also

contain observations, which could provide more information such as whether a term is particular to a certain language variety or region of the world.

The user interface for *TERMIUM Plus* is available in either English or French, and performing a basic search is relatively straightforward.

Try it! What can *TERMIUM Plus* tell you about Canadian usage?

1. Go to the website *for* *TERMIUM Plus*.
2. Locate the search box titled "Which term?"
3. Type or copy/paste the term ***tweet*** and then click on the Launch button or hit the Enter key.
4. Look at the resulting hits. You should see two multilingual search records: one for ***tweet*** (noun) and one for ***tweet*** (verb).
5. Consult the first record. You should see the English information listed first, followed by the information for French. In the French part of the term record, there are some notes about appropriate usage for a Canadian audience.
    - The record indicates that ***gazouillis*** is considered to be the correct equivalent in Canadian French, while the Anglicism ***tweet*** is to be avoided.
    - There is an observation noting that ***gazouillis*** is the term that should be used in Government of Canada documents.
6. Consult the second term record, which focuses on ***tweet*** used as a verb, to learn more about the preferred way of expressing this concept in Canadian French.
    - The correct term to use, both in general and in Government of Canada documents in particular, is ***gazouiller***.
    - The verb ***tweeter*** should be avoided in Canadian French.
7. Conduct another search by entering ***spare tire*** into the search box and then clicking the "Launch" button or hitting the Enter key.

    - What do you learn about the preferences of Canadian English and French from this term record? Although Quebec is the largest French-speaking region in Canada, it's not the only one! Regional preferences can exist in other areas too.

For tips on how to use *TERMIUM Plus* in a more advanced way, consult the FAQ which offers information about how to conduct advanced searches and how to interpret the information displayed on the term records.

Le grand dictionnaire terminologique

Le grand dictionnaire terminologique (GDT) is a term bank that is maintained by the Office québécois de la langue française (OQLF), which is an organization whose mission is to guide and promote the development and use of the French language in the Canadian province of Quebec. For example, if a new concept comes into existence, the OQLF will propose a French-language term to designate that concept.

Like *TERMIUM Plus*, the *GDT* is not a dictionary of everyday words but rather a collection of more

specialized terms, which are drawn from a wide range of subject fields. Given the mission of the OQLF, the term records in the *GDT* contain first and foremost French-language terms, which are often accompanied by equivalents in English and sometimes other languages also, in order to facilitate translation. In addition, term records may contain a definition, an illustration, or notes about usage, including regional preferences.

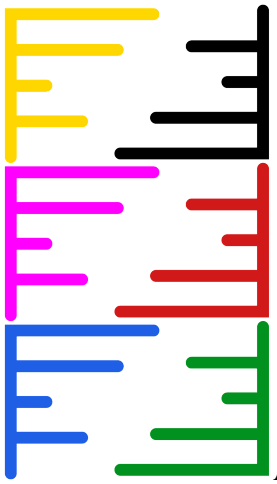
One of the main goals of the GDT is to counsel users about the use of French in Quebec, and so where relevant, the term records are colour-coded with **green**, **yellow** and **red** sections, similar to the colours found on a traffic light. Terms in the **green** section are preferred for use in Quebec, while those in the **yellow** section may be used in some contexts. Meanwhile, terms in the **red** section should be avoided. These guidelines can help translators to select the terms that are appropriate for addressing Canadian audiences.

Try it! What can the *GDT* tell you about Canadian usage?

1. Go to the website for *La Vitrine linguistique* of the OQLF and check the box to search *Le grand dictionnaire terminologique*. Note that the interface for this tool is available only in French.
2. In the search box, type or copy/paste the term ***plugiciel*** and then click on the magnifying glass icon or hit the Enter key.
3. The search should return a list of term records that contain the search term. Click on the first item in the list: **plugiciel FR • plug-in EN**.
4. Consult the French-language portion of the term record, which contains a range of information about which terms are preferred for use, and which are not.
    - The green check mark shows that the French term ***plugiciel*** is recommended for use in Quebec
    - The note explains that in France, the term more commonly used to express this concept is ***module d'extension***.
    - The section marked with a red "do not use" symbol advises users that the Anglicism ***plug-in*** should not be used when writing in French.
5. Consult the English-language section of the term record, which lists the equivalent terms in this language: ***plug-in***, ***plugin***, etc.
6. Scroll to the top of the screen and locate the green rectangle to conduct another search.
7. Type or copy/paste the term ***hockey stick***.
    - What information does the *GDT* provide about the best way to describe this piece of sports equipment in Canadian French?
    - What terms are used language varieties in other parts of the world?

For a more detailed description of the *GDT*, including details on how to conduct advanced searches and how to filter and interpret the results, consult the FAQ (available in French only).

Bilingual Concordancers



Although term banks can be very useful for understanding concepts (e.g. through definitions) or for identifying possible equivalents in another language, sometimes the term or expression that a translator needs may not be included in the term bank. Other times, a translator may want to see a term in context in order to better understand how to use it properly in the target text. In such cases, a free online bilingual concordancer can be very useful.

A bilingual concordancer is a fairly straightforward tool: it allows users to search through a bilingual parallel corpus to find information that might help them to complete a new translation task. For example, if a user encounters a word or expression that they do not know how to translate, they can look in the bilingual parallel corpus to see if this expression has been used before, and if so, how it was dealt with in that previous translation.

In order for a bilingual concordancer to work, the source and target texts must first be aligned, which means that corresponding text segments – usually sentences – are linked together. See the figure below for an example of how the text from the English and French versions of the website of the University of Ottawa's School of Translation and Interpretation can be aligned at the sentence level.

Welcome to the University of Ottawa's School of Translation and Interpretation (STI)	Bienvenue au site de l'École de traduction et d'interprétation (ÉTI) de l'Université d'Ottawa
Translation is one of the most creative and lucrative ways for people who love working with languages to enjoy earning their living.	Quand on aime les langues, traduire représente une activité professionnelle à la fois créative et lucrative.
And in Canada, translation is a constant.	Or, au Canada, la traduction est omniprésente.
The job options are countless: in federal and many provincial governments, in law offices and courts, in NGOs and many other agencies, in film and media, in advertising, journalism, academia, and, of course, in literature.	Les débouchés dans ce domaine sont innombrables : administration fédérale ou provinciale, cabinets d'avocats, palais de justice, ONG, sans oublier l'audiovisuel, la publicité, le journalisme, le monde universitaire et, bien entendu, la littérature.
Since the University of Ottawa is a bilingual institution, classes and exams (and many administrative activities) take place in both languages, and there is a constant interplay between	À l'Université d'Ottawa, institution bilingue, les cours et les examens (tout comme une grande part de l'administration) ont lieu en anglais et en français, ce qui donne lieu à une interaction

Canada's two official languages.	constante entre les deux langues officielles du Canada.

Once the texts are aligned, the translator can use the bilingual concordancer to consult the corpus. By entering a term or expression in the search box, the translator can retrieve all examples of that term from the corpus. The source language sentences containing the search term will be shown on the left, and the corresponding translations in the target language, which should normally contain an equivalent for the search term, will be displayed on the right.

TradooIT Bilingual Concordancer

As part of their TradooIT tool suite, the Canadian company Okidoo Inc. has developed a free online bilingual concordancer that searches a corpus of over 850 million words. The majority of texts the in the corpus come from Canadian government websites, although some texts are taken from international organizations. The source for each sentence is always displayed immediately below the sentence, which allows users to see at a glance whether or not the content is Canadian.

When translating for a Canadian audience, it is preferable to use terms or expressions that are representative of the Canadian variety of English or French, and because TradooIT's corpus contains such a high percentage of Canadian sources, it is easier to identify authentic examples of Canadian language in this resource. There are other similar bilingual concordancers on the web, such as Linguee or Glosbe; however, the corpora used by these other tools contain a greater proportion of non-Canadian texts, meaning that it is more challenging to locate examples of Canadian usage. Therefore, when translating for a Canadian audience, TradooIT's corpus and bilingual concordance are likely to be more helpful.

Try it! How can you use sources to identify authentic Canadian language?

1. Go to the TradooIT bilingual concordancer website.
2. Set the source language to English and the target language to French using the drop-down lists below the search box.
3. Type or copy/paste the term ***parking lot*** into the search box and then click on the button showing a magnifying glass or hit the Enter key.
4. Look at the results, which are displayed in two columns in the centre of the screen labelled English and French.

    - In the English column (on the left), there is a series of English sentences containing the term ***parking lot***.
    - In the French column on the right, there is a series of French sentences that are translations of the English sentences.
        - What French equivalents for ***parking lot*** can you see?
        - In most of the French sentences, the French equivalent is ***stationnement***, but in some cases, ***parking***

5. Look at the source information listed below each sentence. This indicates the text from which each sentence was taken.

    - Notice that most of the sources are from Canadian websites (indicated by a .ca domain name), while a few sources are not Canadian (e.g. sites with a .com domain name).
    - The examples taken from the Canadian sources tend to contain the French Canadian term ***stationnement***, while the examples taken from non-Canadian sources tend to be the ones that use the term ***parking***, which is more common elsewhere.

6. Try doing this same search (for ***parking lot***) in another free online bilingual concordancer tool such as Linguee.
    - What do you notice about the results?
    - You will probably find that the proportion of Canadian sources in Linguee is lower than the proportion of Canadian sources in TradooIT, making it more challenging to identify authentic examples of Canadian usage.
7. Return to the TradooIT bilingual concordancer and try another example.
8. Set the source language to English and the target language to French using the drop-down lists below the search box.
9. Type or copy/paste the term ***octopus*** into the search box and then click on the button showing a magnifying glass or hit the Enter key.
10. Look at the results, taking note of the sentences in the French column on the right-hand side.

    - Some sentences contain the term ***poulpe*** while others contain the term ***pieuvre***.

11. Turn your attention the far left-hand side of the screen and find the subheading "Sources".

    - This list indicates where the texts in the corpus come from.

12. Click on the item called "PARLGCCA", which means that the texts come from the *parl.gc.ca* domain (i.e., texts from the Parliament of Canada). By clicking on "PARLGCCA", only those examples that come from the Parliament of Canada texts are shown.

    - In these Canadian examples, the term ***pieuvre*** seems to be more common.

13. Return to the list of sources and this time click on "EUROPARL", which will show only those examples that come from the European parliament texts.

    - In these European examples, ***poulpe*** seems to be the more commonly used term.

In addition to the basic search and display options, TradooIT's bilingual concordancer has some more advanced features that allow users to incorporate wildcard characters for searching and to apply filters for sorting and viewing the results in different ways. To learn more about the advanced features of TradooIT's bilingual concordancer, consult the online user guide. You may also have noticed that, in addition to displaying the English and French sentences that it retrieves from the corpus, the TradooIT tool also displays any relevant results that it finds from the Canadian term banks that were discussed in the previous section. These term bank results are displayed at the top of the screen above the two columns containing the English and French sentences. Together, this combination of term records and concordance results that come from mainly Canadian sources provides a good indication of the language that is characteristic of the Canadian varieties of English and French.

4

Tools for Investigating Regional Variation in Languages

Sometimes translators may encounter a term and be unsure about whether that term is commonly used in Canada. Other times, translators may find themselves trying to decide between two different terms or expressions that are similar in meaning but may reflect different regional preferences. While the term banks and bilingual concordancers discussed in previous sections can be used to investigate terms, they do not always facilitate quick comparisons of the way that terms are used in different regional varieties of a language. However, there is another type of tool that allows such comparisons to made.

Diatopix

Diatopix is a free online tool that was conceived by Patrick Drouin, a professor at the Université de Montréal, and programmed by Benoit Robichaud. Diatopix allows a user to enter a term or a pair of terms and to see how frequently these terms are used in different regions of the world. For varieties of English, the regions available for comparison are Australia, Canada, Ireland, New Zealand, the United Kingdom (UK), and the United States (US). Meanwhile, the different varieties of French that can be compared are the varieties used in Belgium, Canada, France, Luxemburg, and Switzerland. In addition, Diatopix also works for several varieties of Spanish and Portuguese.

How does it work? As explained on the website, Diatopix allows users to see in a graphical format (e.g. a bar graph or a pie chart) the way that terms or expressions are distributed on the web. To do this, Diatopix uses the Google search engine's custom search and categorizes the results it receives according to the main countries where that language is used (Australia = .au, Canada = .ca, Ireland = .ie, New Zealand = .nz, etc.)

For instance, a user can enter just the term ***wellington boots*** and Diatopix will return an easy-to-read visual display (in the form of a bar chart and a pie graph) showing that this term is not used very often in any region other than New Zealand, and so it would not be a good choice for an audience in Canada.

Similarly, a translator could enter two different terms such as ***sweater*** and ***jumper*** and Diatopix will generate a comparative display showing the relative frequency of the two terms in each of the six regional varieties. In this case, Canadian English shows a preference for ***sweater***, which lines up with the US preference and contrasts with the UK preference.

Diatopix can also be used investigate spelling variants, such as ***analyze*** and ***analyse*** or ***theatre*** and ***theater***. For the first example, Diatopix reveals that Canadian English tends to be similar to US English in preferring the "z", while in the second example, the "re" option is more common in Canadian English, which is similar to the preference in UK English.

Try it!* How can you see if an expression is more common in one language variety than in another?

1. Go to the Diatopix home page and select English from the menu in the upper righthand corner of the screen.
2. Locate the search boxes in the centre of the screen.
3. Type or copy/paste ***toboggan*** into one of the search boxes and then click the "Go" button or hit the Enter key.
4. Consult the results, which are displayed as both a bar graph and a pie chart.

    - Note that this term is used more often in Canada than in any of the other English-speaking regions.

5. Return to the search boxes at the top of this screen and place your cursor in the empty box.
6. Type or copy/paste ***sled*** in the second box and then click the "Go" button or hit the Enter key.

    - What does Diatopix tell you about how the terms ***toboggan*** and ***sled*** are used in different regions of the English-speaking world?
    - Would you feel comfortable opting for ***toboggan*** when writing for a Canadian audience?

7. Compare some spelling variants such as ***labor*** and ***labour*** to see which option is preferred in Canada.
8. Switch the search language to French by choosing Français from the language menu in the upper right of the screen.
9. Type or copy/paste the terms ***bonnet de bain*** and ***casque de bain*** into the two search boxes and click on the "Go" button or hit the Enter key.

    - What do the visual displays suggest about the use of these two options in various regions?

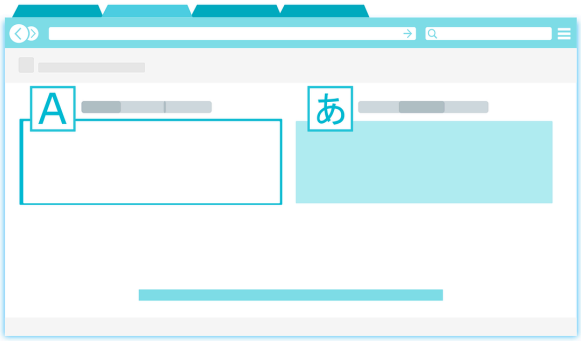
*__Note__: Diatopix uses data from the Internet, which is constantly changing, so the results produced may differ from the results that were shown at the time that this resource was created.

Diatopix does have some limitations that users should keep in mind. For example, this tool cannot distinguish between homographs, which are words that have the same spelling but have different meanings or different parts of speech, such as ***pop***, which can be a noun, verb or adjective. So a search for the difference in regional usage between ***soda*** and ***pop*** will not return meaningful results because ***pop*** has more than one meaning. Likewise, Diatopix cannot handle inflections, so it is necessary to conduct separate searches for a singular and plural form or for conjugated verb forms.

In spite of these limitations, Diatopix still offers helpful food for thought about regional variations. As stated on the Diatopix home page "The tool in itself does not prove anything but it can be used to confirm intuitions regarding word usage or to explore research leads that have to do with lexical variants from a geographical point of view."

5

Machine Translation Tools



Machine translation (MT) software attempts to undertake a translation task from beginning to end. In other words, if a user provides a source text that needs to be translated into another language, the machine translation tool will take that source text and return a translated text in the selected target language. The resulting translation might not be perfect, but the tool will make an attempt and return a completed translation that the user can use as a draft or starting point. In some cases, this draft might be all that is required (e.g. if the user is just trying to get the gist of the message). Other times, the user might need a more polished text, in which case the user may decide to edit the machine translated text in order to improve it. It's important for users to consider the purpose of the translation and to accept responsibility for producing a text that meets expectations, which may mean that it is necessary to edit the initial text produced by the machine translation system.

Can I use MT for courses? It depends! Courses have different learning objectives. In a geography course, demonstrating that you understand the main concepts might be a key objective, so if MT can help you to express those concepts more clearly, that's may be a good use of the tool. But in a language course, one objective might be to show that you have mastered a particular grammatical feature, so the

instructor might want you to demonstrate this without using MT. If you are unsure whether MT is permitted, ask your instructor, and be transparent about using MT (e.g. add a note explaining where and how you used MT).

Because the machine translation system is doing the translation work "behind the scenes", the user interface for a free online machine translation system is usually quite simple. First, a user selects both the source (or start) language and the target (or end) language. Next, the user types or copies the text that needs to be translated into the box under the source language. The machine translation tool will then produce a translation of that text in the target language.

Don't plagiarize! If you use MT to translate a passage from another source that you then integrate into your own work, remember to add a reference to the original material because even though you have changed the language, you still need to cite the original ideas.

Google Translate is probably the best known and most widely used free online machine translation system; however, this system has not (yet) been customized for Canadian English or French. However, there are two other free online machine translation systems – TradooIT Translator and Microsoft Bing Translator – which offer some degree of customization for Canadian language varieties. Before exploring these two tools in more detail, let's first consider some of the basics of how they work in order to better understand how it has been possible to customize these tools for Canadian versions of English and French.

Both TradooIT and Microsoft Bing Translator are examples of *neural* machine translation (NMT) systems. NMT is a very recent approach to machine translation that employs artificial intelligence-based machine learning techniques. The earliest approach to machine translation (dating back to the period after World War II) was known as rule-based machine translation (RBMT) because it tried to get computers to process language in a way that resembled some aspects of how people process language. Therefore, these RBMT systems were supplied with large bilingual dictionaries and an extensive set of grammar rules, and a different dictionary and set of grammar rules had to be created for each language pair. In other words, an RBMT system that was developed to work with English and French could not be adapted easily to work with Arabic and Russian, for example. To enable the system to work with different languages, new dictionaries and grammar rules had to be programmed, which represented a considerable amount of work. What's more, these RBMT systems had limited success because people actually use much more than grammar and vocabulary when communicating. To communicate successfully, people also make extensive use of context and real-world knowledge to interpret phrases that might be ambiguous. However, computers don't have access to real-world knowledge, so they are not well equipped to interpret natural language, which is inherently ambiguous.

Privacy and confidentiality: Information entered into a free online MT tool doesn't disappear when you close the window. Many developers state in their terms of use that they may use it for other purposes (e.g. to fine-tune the tool). Don't enter sensitive information into a free online MT tool. Some clients may not permit translators to use free online tools to translate their proprietary content.

Given the limited success of RBMT, developers decided to try new approach that did not rely on linguistics and world knowledge but which instead allowed computers to do what they are good at: pattern matching and number crunching. This led to the development of *corpus-based approaches*, including the most recent approach known as neural machine translation. In NMT, the developers provide the system with an extremely large bilingual corpus containing source texts and their translations that have been produced previously (usually by professional translators). The computer can

consult the previous translations to find patterns (e.g. X was previously translated by Y ninety-nine times out of one hundred) and "learn" from these examples how to translate new source texts. Of course, this explanation is very simplified, but the essential thing to note is that the machine translation system learns from the very large number of examples that have been provided.

This means that it's much easier to create a tool that can translate different languages because the main thing that is needed is a new bilingual corpus that the system can consult and learn from. By carefully selecting what is included in the corpus, it is possible to teach the system different things, such as a different language variety. Remember, however, that in order to produce high-quality results, it is necessary to have an extremely large corpus (i.e. billions of pages of text), which may be challenging to compile if a particular language or language variety is not widely used.

Let's have a look at some free online NMT systems that have succeeded in using customized corpora of Canadian texts to allow users to generate translations in Canadian English or Canadian French.

TradooIT Translator

The TradooIT tool suite was developed by Simon McDuff, founder of Okidoo Inc. In addition to the bilingual concordancer described in a previous section, the TradooIT tool suite also contains a free online neural machine translation tool. According to the TradooIT website, this neural machine translation system has been trained primarily using Canadian government texts and is thus adapted to the Canadian vocabulary. What's more, this tool works for both Canadian English and Canadian French. Indeed, this tool is dedicated to these particular language varieties and no other language options are available. It is worth noting that since the tool has been trained using mainly government texts, it may not perform as well on other types of texts. Remember that NMT systems learn by looking at the examples in the corpus. Different types of texts have different characteristics, so a machine translation tool that is trained on one type of text (e.g. government texts) may not be well equipped to handle other types of texts (e.g. recipes, textbooks, literature).

Like most free online machine translation tools, the interface for TradooIT Translate is self-evident and the tool is easy to use. Simply paste the source text in the left-hand box and then view the translated text in the right-hand box.

Try it! Can TradooIT Translate produce Canadian language?

1. Go to the TradooIT Translate website.
2. Select French as the source language and English as the target language.
3. Copy/paste the following text into the source text box on the left:

De quelle couleur est leur médaille olympique ? Les plongeuses canadiennes ont gagné la médaille d'argent à l'épreuve féminine du tremplin de 3 *mètres* synchronisé.

4. Look at the proposed English-language target text in the right-hand box.*

- Does the spelling reflect Canadian English's preferred blend of US and UK options?
  - The words ***colour*** and ***metre*** reflect the ***–our*** and ***–re*** endings that Canadian English has inherited from UK English.
  - The term ***synchronized*** reflects the ***–ize*** preference of Canadian English that has been influenced by US English.

***Note**: NMT systems are constantly learning, so the results produced may differ from the results that were shown at the time that this resource was created.

Microsoft Bing Translator

Although Microsoft Bing Translator is not a Canadian product, it does offer an option that is not found in most other well-known free online machine translation tools, which is the option to select "French (Canada)" as a language, rather than only choosing "French". In other words, Microsoft has trained a version of their machine translation system using a corpus of texts that have been written in Canadian French. As a result, the translations that are produced when the "French (Canada)" option is selected will be in a version of French that is more typically used in Canada.

Try it! Can Microsoft Bing Translator produce Canadian French?

1. Go to the Microsoft Bing Translator website.
2. Select English as the source language and French (Canada) as the target language.
3. Copy/paste the following text into the source text box on the left:

Soccer is a very popular sport. Sometimes kids even play in empty parking lots.

4. Choose "French" as the target language
5. Look at the proposed target text in the right-hand box.*
   - Notice that ***soccer*** has been translated as ***football***,
   - See that ***parking lots*** has been translated as ***parkings***.
6. Change the target language to French (Canada) and watch how the translation changes.
   - Now ***soccer*** has been translated as ***soccer***.
   - This time, ***parking lots*** has been translated as ***stationnements***.

***Note**: NMT systems are constantly learning, so the results produced may differ from the results that were shown at the time that this resource was created.

Unfortunately, Microsoft Bing Translator does not have a comparable option for Canadian English. Instead, "English" is the only option and there are currently no regional varieties available.

6

ChatGPT

Chat Generative Pre-trained Transformer – better known as ChatGPT – is an artificial intelligence tool developed by OpenAI. ChatGPT is based on a large language model (LLM). The version of ChatGPT that is based on the model GPT-3.5 is available online free of charge. As a chatbot, ChatGPT is capable of generating text in response to instructions or questions – known as **prompts** – that are provided by human users. Prompt engineering is the process of structuring the question or instruction in a clear and precise way that makes it easy for a generative AI tool such as ChatGPT to interpret and to carry out the task (e.g., answering the question or following the instruction).

Large Language Models (LLMs)

Like the machine translation systems discussed in the previous section, ChatGPT is a data-driven tool. In the case of ChatGPT, it is based on an LLM, which includes a very, very large corpus of texts that is used to train the tool. By looking at the examples in the corpus, the tool learns what texts look like, and based on these examples, the tool learns to generate new texts. It's important for you to recognize that ChatGPT and other LLM-based tools do not understand the *meaning* of the text that they are working on. Rather, LLMs produce text based on the probabilities of which words follow the words that it has already seen or generated. The result is a text that looks good, but its meaning may be strange because the tool does not actually understand what it has written.

ChatGPT and translation

Unlike the machine translation tools mentioned in the previous section on machine translation (i.e., TradooIT Translator, Microsoft Bing Translator), ChatGPT was not designed specifically to translate texts. Rather, ChatGPT can be used for a wider variety of language-related activities, including answering questions, summarizing long texts, simplifying specialized texts, and generating new texts. Even though it was not designed as an automatic translator, ChatGPT can also be used to translate texts. However, as is the case with machine translation tools, the results are unlikely to be perfect and it is important to verify the translated content before sharing it. One interesting aspect of this tool,

particularly for anyone who is translating for a Canadian audience, is that you can specify that you want to translate into a *Canadian* variety of English or French.

Try it! Can ChatGPT produce Canadian language?

1. Go to the ChatGPT website. If you have never used ChatGPT before, follow the instructions to register for a **free** account. When you first access the site, a window will open to give you some tips. You can close the window by clicking on the button that says "OK, LET'S GO".
2. In the ChatGPT message box at the bottom of the screen, enter the following prompt:
    - "**Translate the following sentence into *English:* L'adresse de l'ancien théâtre ne figure pas dans le catalogue.**"
3. Look at the proposed English-language target text.* Does the spelling of words such as ***theatre*** and ***catalogue*** reflect Canadian English, or does it appear in another variety of English?
4. Try again by entering the following more specific prompt in the ChatGPT message box:
    - "**Translate the following sentence into *Canadian English:* L'adresse de l'ancien théâtre ne figure pas dans le catalogue.**"
5. Look at the output now. Has the spelling of ***theatre*** and ***catalogue*** been changed to align with Canadian spelling?
6. Now try translating into French by entering the following prompt:
    - "**Translate the following sentence into *French:* Marie had blueberry pancakes for breakfast.**"
7. Look at the proposed French-language target text.* Take note of the translations for the words ***blueberry*** and ***breakfast*** in particular. What variety of French do these translations reflect?
8. Now adjust the prompt to tell ChatGPT to produce Canadian French specifically:
    - "**Translate the following sentence into *Canadian French:* Marie had blueberry pancakes for breakfast.**"
9. Look again at the translations for ***blueberry*** and ***breakfast***. Have these been changed to terms that are more typical of Canadian French?

***Note**: NMT systems are constantly learning, so the results produced may differ from the results that were shown at the time that this resource was created.

Although the focus of this OER is on translation, you can use ChatGPT to help you "Canadianize" other types of text too. For instance, if a text has been written in American or British English, you can ask ChatGPT to rewrite it in Canadian English. Likewise, if a text has been written in a European variety of French, you can ask this tool to transform it into Canadian French.

Language Portals

A portal is a specially designed web-based platform that collects information on a given topic from different sources and presents it in a uniform way, often with a common user interface.

Language Portal of Canada

The Language Portal of Canada is maintained by the Government of Canada's Translation Bureau. The resources found in the portal are intended for all Canadians who want to improve their language skills and so this diverse collection includes information about language learning, quizzes to test language skills, basic and expert advice on improving writing, a blog written by and for language lovers, and the *TERMIUM Plus* term bank that was discussed in detail in a previous section. The portal has an Official Languages Hub® dedicated to resources for English and French, as well as a section containing learning and teaching resources for some of Canada's Indigenous languages.

As described on the portal's website, the Language Navigator is a useful search tool for Canadians who want to improve their skills in Canada's official languages (English and French). Indeed, Language Navigator has incorporated content of *The Canadian Style*, a guide that gives concise answers to questions concerning written English in the Canadian context, as well as *Le guide du rédacteur,* which does the same for Canadian French.

Using the site's Official Languages Hub® search tool, which permits searching by keyword and theme, users can simultaneously search a wide variety of official languages resources from various Government of Canada departments and agencies. The Hub can be used to locate standards and policies, promotion tools, best practices, learning resources, and other useful information on numerous aspects of official languages in Canada. As such, the Language Portal of Canada is a useful resource to consult when translating specifically for a Canadian audience.

OQLF's La Vitrine linguistique

The *Banque de dépannage linguistique* (BDL) is a French-language resource maintained by the *Office québécois de la langue française* (OQLF) as part of *La Vitrine Linguistique*. Here, users can find advice about grammar, spelling, punctuation, acronyms, pronunciations, writing and more – all with a special focus on Canadian French. Alongside the BDL and the *Grand dictionnaire terminologique* term bank that was described in a previous section, the OQLF's *La Vitrine linguistique* portal offers other tools to

support language learners, templates for common types of texts, a database of place names in the province of Quebec, and more.

1

Concluding Remarks

When translating for a Canadian audience, it makes sense to address readers in the language of their region, which is Canadian English or Canadian French. This Open Educational Resource has provided an overview of a range of tools and resources that can help translators to make informed choices when translating for Canadian audiences. It's not a comprehensive guide, but it's enough to get you started, so put your ***runners*** on and get going!

2

Acknowledgements

This Open Educational Resource was initially produced with a grant from the University of Ottawa Library, and it was updated (version 2) with support from eCampusOntario. Special thanks to librarian Mélanie Brunet (ORCID iD) for support throughout the production process.

Copyright for images:

- Cover art and end-of-chapter images adapted from Maple Leaf by elgraphics on Pixabay, Pixabay License.
- Art for practical exercises from Let's do it by Maklay62 on Pixabay, Pixabay License.
- Introduction art from Welcome by Tumisu on Pixabay, Pixabay License.
- Chapter 1 art from Globe by geralt on Pixabay, Pixabay License.
- Chapter 2 art from Network by geralt on Pixabay, Pixabay License.
- Chapter 3 art adapted from The Graph by RoadLight on Pixabay, Pixabay License.
- Chapter 4 art from Countries by OpenClipart-Vectors and Flip Chart by janjf93 on Pixabay, Pixabay License.

- Chapter 5 art from Translate by 200degrees on Pixabay, Pixabay License.
- Chapter 6 art from AI Generated ChatGPT by Franz26 on Pixabay, Pixabay License.
- Chapter 7 art from Student by StartupStockPhotos on Pixabay, Pixabay License.
- Conclusion art from Feet by MabelAmber on Pixabay, Pixabay License.

About the Author

Lynne Bowker, PhD, CTran, FRSC (ORCID iD)

Full Professor

School of Translation and Interpretation

University of Ottawa, Canada

Email: lbowker@uottawa.ca

Twitter: @bowkerl

Website: https://uniweb.uottawa.ca/members/415