

ABSTRACT

USING SEMANTIC AND PRAGMATIC INFORMATION TO DETERMINE SYNTACTIC DECISIONS IN TRANSLATION OF THE PASSIVE VOICE FROM ENGLISH TO FRENCH

Ashli Fain, M.S.
Department of Computer Science
Northern Illinois University, 2018
Dr. Reva Freedman, Director

I investigated statistically significant factors that effect syntactic changes in French and English during translation of *by* phrases using the passive voice. Relative clauses containing this structure were also looked at. This feature was chosen from information that was gathered from experts in the form of academic texts on English-French translation as well as handbooks intended for an audience of professional translators. Data came from the proceedings of the Canadian Parliament (the Hansard), which are published in French and English. Passive sentences were collected from the output of the Stanford Parser. Animacy was found to be a significant factor in whether or not the arguments of a passive *by* phrase would change order in translation. Information status will need further investigation. While relative clauses are still subject to this effect, this syntactic shift occurs far less often than in sentences not containing relative clauses.

NORTHERN ILLINOIS UNIVERSITY
DE KALB, ILLINOIS

DECEMBER 2018

USING SEMANTIC AND PRAGMATIC INFORMATION TO DETERMINE
SYNTACTIC DECISIONS IN TRANSLATION OF THE PASSIVE VOICE
FROM ENGLISH TO FRENCH

BY

ASHLI FAIN
©2018 Ashli Fain

A THESIS SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE
MASTER OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

Thesis Director:
Dr. Reva Freedman

ProQuest Number: 10977987

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10977987

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

TABLE OF CONTENTS

	Page
LIST OF TABLES	iv
LIST OF FIGURES	v
Chapter	
INTRODUCTION.....	1
Hypotheses.....	2
BACKGROUND	3
Pragmatics.....	3
Syntax.....	9
Relative Clauses	10
Semantics	12
Animacy	14
APPROACHES TO TRANSLATION.....	16
METHODOLOGY	19
DATA COLLECTION.....	22
Passive Sentences	23
Noun-Phrase Arguments.....	27
Information Status.....	29
Limitations.....	31
Animacy	34
<i>Tour de Présentation.....</i>	34

Limitations	36
RESULTS	38
Other	40
SCOPE	43
FUTURE WORK	44
REFERENCES	45
Speeches	47

LIST OF TABLES

Table	Page
1. Information Status.....	7
2. Semantic Roles.....	14
3. Part of Speech Tags	25
4. Passive Tags and Examples of Matching Text	25
5. Skipped Tags	26
6. Example of Skipped Text.....	26
7. Discourse Status (Based on Birner, 2013).....	33
8. Related Words	37
9. Animacy of Arguments Compared.....	38
10. Passive Verbs Replaced Using a <i>Tour de Présentation</i>	38
11. <i>By</i> Phrases in Relative Clauses vs Regular Clauses.....	40

LIST OF FIGURES

Figure	Page
1. Processing the Hansard Text	23
2. CoreNLP Output	24
3. Partial Row from Database	24
4. CoreNLP Tree Printed from NLTK.....	27
5. Relative Clause.....	28
6. SBAR Phrase	29

INTRODUCTION

I investigated statistically significant factors that effect syntactic changes in French and English during the translation of *by* phrases using the passive voice. Relative clauses containing this structure were also looked at. This feature was chosen from information that was gathered from experts in the form of academic texts on English-French translation as well as handbooks intended for an audience of professional translators. Data came from the proceedings of the Canadian Parliament (the Hansard), which are published in French and English. Passive sentences were collected from the output of the Stanford Parser. Animacy was found to be a significant factor in whether or not the arguments of a passive *by* phrase would change order in translation. Information status will need further investigation. While relative clauses are still subject to this effect, this syntactic shift occurs far less often than in sentences not containing relative clauses.

The explanations given by experts for the avoidance of translating the passive voice as such in French are related to syntactic, semantic, and pragmatic factors. Syntax-level factors concern the structure of the language. Semantic factors concern the meanings of words, phrases, or text. Pragmatic factors are the meanings of utterances in context.

Using these factors as a foundation for this study, I produced a computational treatment of what effects syntactic changes in the process of translation. Syntactic analysis of the Hansard text was obtained from Stanford CoreNLP's constituency parser. Data was collected by a Python program that I wrote and stored in a database implemented as a .csv file.

Hypotheses

1. A *tour de présentation* will not be used to move a less animate noun phrase in front of a more animate one.
2. A *tour de présentation* will be used significantly less often to translate *by* phrases in relative clauses than *by* phrases that do not happen within a relative clause.
3. Where one of the arguments of a passive verb is deleted, inanimate noun phrases will be deleted more often than animate noun phrases.

BACKGROUND

Factors that were studied included pragmatics, syntax and semantics. Syntax concerns the structure of the language, semantics concerns the meanings of words and phrases, and pragmatics concerns utterances in context.

Pragmatics

A sentence consists of a subject and a predicate. A proposition is something, either true or false, that can be represented as a sentence. Sentences are paraphrases if they express the same proposition.¹ Sentences (1) and (2) are paraphrases:

- (1) Sam loves Jamie.
- (2) Jamie is loved by Sam.

The same proposition may be expressed by several sentences in English. Similarly, in the translation of a sentence, translators have some freedom with word order (Darbelnet and Vinay, 1958). Sentences (4) and (5) are both acceptable translations of (3). Sentence (4) uses the passive voice like the original. In (5) the noun phrases have changed places, and the sentence is active.

- (3) The bill was approved by the Senate.
- (4) Le projet de loi a été approuvé par le sénat.
The bill was approved by the Senate.
- (5) Le sénat a approuvé le projet de loi.
The Senate approved the bill.

¹ Betty Birner (Department of English, NIU), personal correspondence.

The experts provide explanations going beyond syntactic constraints as to why translators may choose certain sentence constructions over others.

One factor consists of how certain pragmatic and semantic content determines the thematic structure of the sentence. Most modern pragmatic theories concern the interactions between a speaker, the entity producing the utterance, and the hearer, the entity or entities to whom the utterance is directed. The pragmatic content described by the experts includes information status, whether the information is known or unknown to the hearer, as well as importance to the speaker and what is useful to the hearer. The speaker in the Hansard is whoever is currently holding the floor; the hearers are the rest of those present in the House of Commons not currently holding the floor.

Darbelnet and Vinay (1958) use the distinction that Bally (1941) and Blinkenberg (1928) made between a psychological or thematic order and the grammatical order of a sentence. The concept of a psychological order and its interactions with grammatical order was studied in depth by Andreas Blinkenberg. Blinkenberg drew a distinction between the grammatical order, the order imposed on words by the structure of the language, and the psychological order, the order imposed by the thought process of the speaker.

According to Blinkenberg, the psychological subject A and predicate Z most often correspond with the grammatical subject and predicate, but they can also be in opposition to one another. Blinkenberg states that A usually contains known or old information and Z usually contains unknown or new information. Darbelnet and Vinay (1958) state that both French and English have a tendency to place A first in general. However, if Z is placed first in the English sentence, it can be necessary for translators to reconfigure it in French to reflect a preference for the order AZ. This can be done using a *tour de présentation* to rearrange the structure of the

sentence. In (6), the text contains clauses that have a causal relationship, and according to the authors, the *tour de présentation* is used to place the cause before the effect.

- (6) F: [Prévenus à temps,] [ils purent rebrousser chemin]
 [avant d'être surpris pas l'orage.]
 [Warned in time,] [they could turn back] [before being
 surprised by the storm.]
 E: [They were able to turn back] [before the storm
 overtook them,] [as they had been warned in time.]

Bally claims that if all entities in the phrase are of the same information status, being equally more or less known to the hearer, the grammatical order is maintained. He presents the following examples where dislocation is used to convey spoken emphasis in order to demonstrate how A and Z can be ordered:

Order = AZ

[Cette lettre,]_A [je ne l' ai jamais reçue.]_Z
 [This letter,]_A [I it have never received.]_Z
 'I never received **this letter.**'

Order = ZA

[Je ne l' ai jamais reçue,]_Z [cette lettre.]_A
 [I it have never received,]_Z [this letter.]_A
 '**I never received** this letter.'

Additionally, he notes that each part of these parts of a sentence can be further broken down into A and Z.

Order = AZ (where Z = AZ)

[Cet homme,]_A [[moi,]_(A) [je ne le connais pas.]_(Z)]_Z
 [This man,]_A [[me,]_(A) [I him know not.]_(Z)]_Z
 'I don't know **this man.**'

Order = AZ (where Z= ZA)

[Cette homme,]_A [[je ne le connais pas,]_(Z) [moi.]_(A)]_Z
 [This man,]_A [[I him know not,]_(Z) [me.]_(A)]_Z
 'I **don't know this man.**'

Bally expands on the idea of the psychological subject A as what is important to the speaker; the psychological predicate Z is what the speaker wants the hearer to know about the subject.

The experts state that information being old/known or new/unknown to the hearer is a factor that determines the structure of a translated sentence. Each of the authors ascribe slightly different characteristics to the psychological order consisting of A and Z. All of the characteristics attributed to A and Z are determined from the context of the utterance; therefore, it is pragmatic, and can be examined with other pragmatic frameworks, such as Ellen Prince's description of information status.

Ellen Prince (1992) offers a more detailed taxonomy of the information status of discourse entities than the known/unknown values ascribed to A and Z previously discussed. Entities may be hearer-old or hearer-new according to the speaker's beliefs about what the hearer already knows. In (7a), the speaker assumes that the hearer already knows of an entity named *Sandy Thompson*; in (7b) the speaker believes that the hearer does not, and so introduces a new entity.

(7a) I'm waiting for it to be noon so I can call Sandy
Thompson.

(7b) I'm waiting for it to be noon so I can call someone
in California.

An entity may also be new or old depending on whether or not it was previously evoked in the discourse. Sandy Thompson becomes discourse-old in (8b), which is a response to (8a):

(8a) I'm waiting for it to be noon so I can get in touch with Sandy Thompson.

(8b) Why are you trying to get in touch with Sandy Thompson?

Entities may be categorized according to Table 1. Descriptions of these categories are as follows:

- **Brand new:** the entity is assumed by the speaker to be unknown to the hearer, and it has not been mentioned in the discourse.
- **Unused:** the entity is assumed by the speaker to be known to the hearer, and it hasn't yet been mentioned in the discourse.
- **Evoked:** the entity is assumed by the speaker to be known to the hearer, and the entity has been previously mentioned in the discourse.

However, an entity cannot be discourse-old and hearer-new. An entity can't be evoked in the discourse without also then being known to the hearer.

Table 1, Information Status (based on Prince, 1992)

	Discourse-new	Discourse-old
Hearer-new	Brand new	N/A
Hearer-old	Unused	Evoked

Information that can be inferred from the surrounding discourse is treated as discourse-old as well, even if it hasn't been explicitly evoked (Birner, 2013). Within Prince's framework, this information is considered *inferrable* (Prince, 1992). For example, if I say, "I walked into the classroom and opened the window," *the window* is not discourse-new, but rather inferrable, as it's known that classrooms can have windows.

The information status of discourse entities has implications for the passive voice. Passive *by* phrases require that the information in the preposed constituent not be any less familiar than that in the postposed constituent (Birner, 2013). In (9), *competence* has been previously mentioned in the discourse, whereas *many senior officials* is being evoked for the first time.

- (9) There is no easy way to competence in a second language, but in three and a half years, *enough such competence has been acquired by many senior officials...*
(Trudeau, 1968)

Ten is the beginning of an article, and neither noun phrase represents information that has already been mentioned in the discourse. However, since they are of similar information statuses, the passivization is acceptable.

- (10) *An eye-popping \$10 billion in long-term aid for Haiti was pledged by 48 countries and international institutions...*
("Quick Updates," Christian Science Monitor, qtd. in Birner, 2013)

Therefore, Blinkenberg's claims about the information status of known and less known corresponding to psychological A and Z, while applicable to subjects and predicates, may not be applicable to the choice of noun-phrase word order in translation of the passive voice. This is because a noun-phrase representing less familiar information would not be preposed in a passive sentence in the first place.

- (11a) I walked into the classroom and **the windows** had been opened by *the students*.

- (11b) ?I walked into the classroom and **the refrigerator** had been opened by *the students*.

However, if these claims hold, then it would be constituents of similar information statuses that would allow translators more freedom with word order. In order to reflect a preference for the order AZ, constituents in an English sentence that already honor the

preference for more familiar information to be first in passive would remain as such in French, rather than being translated using a *tour de présentation*.

Syntax

While the passive voice exists in French, it is used with less frequency than in English. According to a study by Soudieux (Delisle, 2003), 71.3% of passive verbs in English texts became active in French texts. According to Ginette (Delisle, 2003), who produced similar statistics, this is due to structural differences in the languages. While languages such as French and German provide many ways to produce the impersonal tone used in technical writing, English has only the passive voice as a way of not directly naming the agent of an action.

Additionally, in English, a verb doesn't need to be transitive, i.e., taking a direct or indirect object as an argument, in order to be passivized (Darbelnet and Vinay, 1958). For example, "The issue was argued about." The authors also give numerous higher-level differences between the languages that can contribute to a change in sentence structure that have not been explored statistically.

Passive and active predicates can be argued to have the same deep structure (Radford, 2009). According to Radford, evidence for this is demonstrated in the semantic roles of *by* phrases. Semantic roles describe the way in which the arguments of a given predicate participate in the action that it denotes. The complement of *by* has the same semantic role as the corresponding subject in an active sentence. Additionally, the complement of the active verb serves the same function as the passive subject. For example, *x* is the agent in both (12a) and (12b).

- (12a) x attacked y.
 (12b) y was attacked by x.

This is possible because it is reasonable to assume that the arguments of a predicate receive their semantic roles in a uniform manner, and therefore originate in the same position in the deep structure. This is described in the following principle of Universal Grammar:

Uniform Theta Assignment Hypothesis/ UTAH
 Constituents which fulfill the same thematic role with respect to a given predicate occupy the same initial position in syntax. (Radford, 2009)

Relative Clauses

Passive *by* phrases also occur within object relative clauses. Object relative clauses are those which contain a relative pronoun that functions as the object of the verb (Aygen, 2014).

They modify the antecedent noun-phrase in the higher clause (Radford, 2009). For example:

- (13) The conservative government... brought in a **computer program** *that was developed by an outside private contractor*. (Don Davies, 2018)

The relative clause *that was developed by an outside contractor* modifies *a computer program*.

Given that the noun-phrase in a relative clause is part of a modifier for a noun phrase in the higher clause, a possibility is that the word order of these types of clauses will not be as prone to the types of syntactic shifts seen in other *by* phrases.

Because examples have been found in the English to French translations in the corpus where a *tour de présentation* was used to translate a relative clause containing a *by* phrase, they are being taken into account. Several shifts occur in (14):

(14) [**Bill C-59**,] [which had been introduced by *the previous Conservative government*], [had provisions in it that took a heavy-handed approach..].

[*Le gouvernement conservateur* avait présenté **le projet de loi C-59**] [qui prévoyait notamment une approche musclée...]

[*The Conservative government* had introduced **Bill C-59**], [which provided a heavy-handed approach...]
(Mark Gerretsen, 2018)

The conservative government, the noun phrase contained in the lower clause, switches places with *Bill C-59*, the noun phrase in the higher clause, which the relative pronoun *which* refers to. The third clause in the English sentence also contains a relative clause: ...*had provisions in it that [took a heavy-handed approach]*. In the English text, *a heavy-handed approach* modifies *provisions*. In the newly flipped sentence in the French translation, the relative pronoun *which* now allows *a heavy-handed approach* to modify *Bill C-59* rather than the fact of it having been introduced by the previous Conservative government. However, the sentence still expresses the same proposition, as well as the satisfaction of the syntactic features required by the relative clause. However, it appears more difficult to rearrange such a sentence without altering its meaning or rendering it ungrammatical.

Regular passive sentences have greater freedom with word order. The following sentences are paraphrases, and the complement of the *by* phrase, *a contractor*, has the same thematic role as the subject of the active sentence.

A computer program was written by *a contractor*.
A contractor wrote **a computer program**.

In a relative clause, the *by* phrase within the lower clause may be changed from passive to active and maintain a representation of the same proposition:

They brought in **a computer program** that was written by
a contractor.

They brought in a computer program that a contractor wrote.

However, it is difficult to use a *tour de présentation* without changing the meaning of the sentence more or less dramatically, or rendering it ungrammatical:

They brought in a contractor who wrote a computer program.

?They brought in a contractor that a computer program wrote.

*They brought in a contractor a computer program that wrote.

Regular *by* phrases contain two noun phrases linked by the preposition *by* and a passive verb. Object relative clauses may also contain a *by* phrase and passive verb; however, the second noun-phrase argument is contained in the lower clause that modifies the noun phrase in the higher clause. The deep structure of relative clauses will also be different, with the relative pronoun originating as the complement of the verb instead of the passive subject (Radford, 2009).

Because they are possibly subject to the same syntactic shifts that can occur in translation, they are also being looked at in this study. Because of these syntactic differences, they require their own category separate from other *by* phrases.

Semantics

Experts claim that the French language has a preference to name the agent of the action and to place that agent at the head of a sentence (Darbelnet and Vinay, 1958). In (15), a *tour de présentation* is used to place the animate NP, thematic agent *les témoins*, at the front of the sentence, ahead of the inanimate theme, *the amendment*.

(15) E: [**The amendment** was also commented on favourably by *the witnesses who appeared before the committee to discuss the bill.*]

F: [*Les témoins qui ont comparu devant le comité pour discuter du projet de loi* étaient eux aussi en faveur de **cette modification.**]

[*The witnesses who appeared before the committee to discuss the bill* were also in favor of **this amendment.**]

(Goguen, 2013)

The subjects of active verbs and the complements of passive verbs serve the same thematic function and are subject to the same pragmatic restrictions on choice of arguments, which would include animacy (Radford, 2009).

Radford (2009) states that animate participants of an action can occur in any order and still have the same role. According to Fillmore (1968), the ability of an entity to fill the role of the semantic agent, the entity responsible for an action, is linked with animacy. Additionally, actions associated with semantic agency imply deliberateness. However, there are exceptions, such as in idiomatic uses of a verb. For example, the agent and the patient of the verb “kill” must inherently be animate. However, it can also be used idiomatically where neither agent nor patient are animate. For example, “Bill C-4 was established... to kill two pieces of legislation.”

Fillmore describes a noun phrase that has been assigned the dative case as the animate entity being affected by the state or action of the verb. In more recent literature, this is termed the *Experiencer* role (Radford, 2009 etc.) Thematic roles relevant to this study are described by Radford as in Table 2 as follows:

Table 2, Semantic Roles (Radford, 2009)

Role	Description
Theme	Entity undergoing the effect of some action
Agent	Entity instigating some action
Experiencer	Entity experiencing some psychological state

According to Radford, thematic structure has been argued to play an important role in a wide range of other phenomena, which includes the syntax of the passive voice.

Canadians are shocked by the reports.
 [experiencer] [theme]

The minister was appointed by the board.
 [theme] [agent]

Animacy

Semantic factors noted by experts include a preference in French to place more animate noun phrases in front of less animate noun phrases (Darbelnet and Vinay, 1958). Both Fillmore (1968) and Levin (1993) define human institutions – organizations or institutions made up of human participants – as animate. This is the definition of animacy that was used in this study. Examples of animate noun phrases in the corpus include individuals, government organizations, and other groups of people. Inanimate noun phrases include ideas, amendments, bills, and actions.

There are many examples where an agent in the post-verbal position in a passive sentence remains in place in translation. In (16), the passive voice is conserved from English to French even though there is an agent present, supporting the idea that a *tour de présentation* doesn't happen simply to place the agent first.

(16) E: [Mr. Speaker, earlier this month...] [**the safety of Christians** was yet again threatened by *those pursuing a path of religious intolerance.*]

F: [Monsieur le Président, le début du mois...] [**les chrétiens** ont nouveau été menacés par *des adeptes de l'intolérance religieuse.*]

[Mr. Speaker, earlier this month...] [**Christians** were yet again threatened by *those pursuing a path of religious intolerance.*] (Sgro, 2011)

What is notable are the changes made to the subject noun phrases. Although *des adeptes de l'intolérance religieuse*, being animate and being the thematic agent, would make an acceptable candidate for a *tour de présentation*, that option is not chosen. Instead, *the safety of the Christians* is shortened to *les chrétiens*, referring to a concrete group of people rather than the abstract concept of their safety while maintaining the idea that an action was being undergone. This change elevates the theme (*les chrétiens*) to the level of animacy of the agent.

There appears to be a strong preference for moving animate noun phrases to the front of the sentence in French translation, but this also is not always enough to motivate a *tour de présentation*. There are numerous examples of the passive voice being translated as is regardless of the presence of an animate noun phrase.

APPROACHES TO TRANSLATION

Textbooks on translation (e.g., Darbelnet and Vinay, 1958, Delisle, 2003) give syntactic alternatives that can be used in French to replace the English passive. Experts on translation suggest many options for translating the passive voice from English into French.

One option is to change the point of view of the sentence from passive to active by using a *tour de présentation*. In (17), Pierre Trudeau has previously discussed the presence of two languages in Quebec:

(17) E: [*In the past, this underlying reality of our country has not been adequately reflected in many of our public institutions.*]

F: [Dans le passé, **nos institutions publiques** n'ont pas reflété adéquatement *cette réalité,*] [*qui est à la base même de notre pays.*]

[In the past, **our public institutions** have not adequately reflected *this reality,*] [*which is the very basis of our country.*] (Trudeau, 1968)

Using one tactic suggested by authors to avoid the passive voice, the point of view is changed. *Nos institutions publiques* moves to the front of the clause, and *cette réalité* is relocated to the end.

An entire additional text unit was created in order to elaborate on *cette réalité*. At times, it is necessary to expand upon a word or clause with additional words to create a more precise meaning, referred to as *étoffement*. The opposite can be done for similar reasons, where words are removed, known as *effacement*. In the above example, *cette réalité* refers to something

previously explicitly evoked in the discourse, and *nos institutions publiques* is inferrable. *Nos institutions publiques* is more animate than *cette réalité*.

The reflexive form can be used for middle-voice verbs, where the subject is not responsible for the action:

(18) E: [Surely these arguments are based on fear...]

F: [Il est certain que ces arguments se fondent sur la peur...] (Trudeau, 1968)

In instances where the subject is animate and human, the indefinite pronoun *on*, which translates to *one* or *we*, is used (Grellet, 1985) :

(19) E: [This is easy to state,] [and it has been repeated in hundreds of patriotic speeches...]

F: [Cela est facile à dire,] [et on l'a répété dans des centaines de discours patriotiques...] (Trudeau, 1968)

It is often claimed that French uses the passive voice far less frequently than in English. However, the passive voice can be conserved from English to French, with or without the presence of a semantic agent, and often is:

(20) E: [A program of language training for federal public servants was started in 1964 and has since been greatly expanded...]

F: [Un programme de formation linguistique destiné aux fonctionnaires fédéraux a été institué en 1964 et a été fort élargi...] (Trudeau, 1968)

Additionally, active sentences in English can be translated using the passive voice in French.

Whereas the other options to translate the passive voice mean selecting a new verb or inserting a reflexive or indefinite pronoun, several syntactic changes take place to translate phrases using a *tour de présentation*. Either the active form of the verb is used in order to change

the point of view or a new verb is chosen that allows the agent of the passive sentence to become the subject of the newly-translated active sentence. The theme or patient also becomes the object of the translated sentence. It's reasonable to believe that this method of translation is applied according to certain constraints, rather than randomly.

METHODOLOGY

The basis for differences to search for in the corpus was formed by gathering information from experts on translation. Syntactic analysis of the Hansard text was obtained with Stanford's coreNLP constituency parser (Manning et al., 2014). Data was collected and recorded in a database by a program that I wrote in Python and analyzed using the χ^2 test.

1. I gathered information from books intended for an audience of professional translators about what factors have potential to effect a syntactic shift. Many of these handbooks cite academic texts that seek to place translation within a linguistic framework. *Comparative Stylistics of French and English* (Darbelnet and Vinay, 1958) is one such academic text. Darbelnet and Vinay draw on pragmatic ideas from earlier works by Blinkenberg (1928) and Bally (1941). From these concepts I formed the basis for the features to look for in the corpus.

Many explanations are given by experts as to why the passive voice often poses a challenge for translators, as well as several methods for translating it. The explanations given for the avoidance of translating the passive voice as is in French are related to syntactic, semantic, and discourse-level factors. Syntactic factors are those concerning the structure of the language; semantic factors include animacy and the thematic roles present in the sentence, which are tied to animacy. Pragmatic information includes

information status, in particular whether the speaker believes that the information represented by the noun phrase is known or unknown to the hearer.

The *tour de présentation* in particular showed a potential to have an effect on syntax, as it is a more costly operation than simply maintaining the word order and replacing the passive verb with an appropriate active one.

2. I chose to focus on uses of the *tour de présentation*. I focused on the passive voice where two entities are present connected by the preposition *by* because the *tour de présentation* requires that there be two or more candidates for reordering in the phrase. I observed where the use of the *tour de présentation* occurred in the corpus. While it is used outside of the passive voice, it appears frequently to translate passive *by* phrases.
3. I determined what constraints may be factors in where translators choose to use a *tour de présentation*. I looked at the qualities of the noun phrases involved in passive sentences, found by using regular expressions. These qualities included the semantic roles and information status of the noun phrases. Certain examples where the word order was preserved that showed changes to the characteristics of the noun phrases – one noun-phrase argument being made more or less animate to become equal with the second one – showed that animacy was also a potential factor. Animacy is frequently discussed in the literature as a possible factor that affects word order. Animacy was investigated as well.
4. Text was cleaned up for processing and labelled to ensure proper alignment. It was then parsed with Stanford's coreNLP constituency parser. The Hansard corpus is not aligned by sentence but rather by paragraphs delimited by new lines. The punctuation that the parser uses to delimit text for outputting parse trees was not always used in the same place across translations and was not a reliable marker to align the text by. Using parser

settings to group trees according to new lines doesn't solve the problem of comparing the correct parallel sentences, and this issue was dealt with later.

5. I wrote a program to extract and process text spans containing passive structures to compare them for the selected constraints that may be factors in syntactic shifts.
6. I built a .csv file and ran statistical tests on the data.

DATA COLLECTION

First, the passive sentences were detected. Then, the noun-phrase arguments were extracted. Information was collected and recorded in a database implemented as a .csv file about those noun phrases, including their information statuses, animacy, and determining whether a *tour de présentation* was used. Figure 1 shows the pipeline for the process of collecting data from the Hansard text.

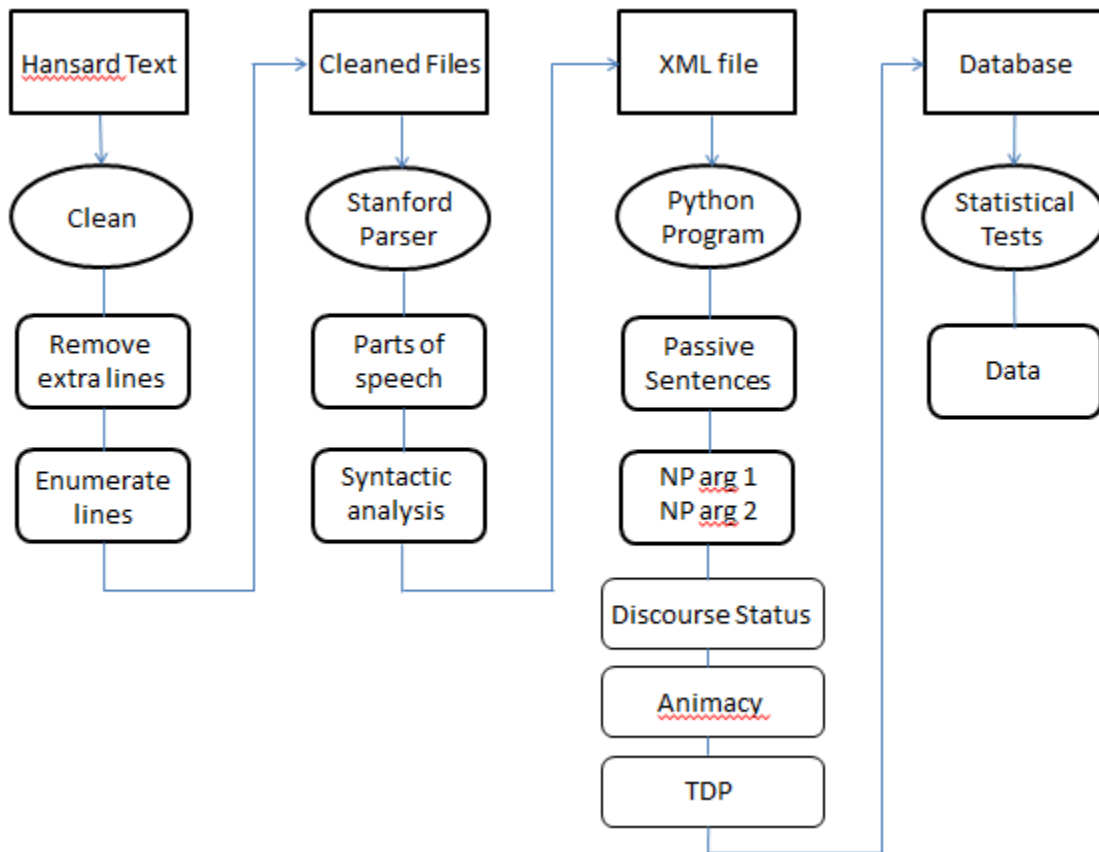


Figure 1, Processing the Hansard text.

Passive Sentences

Stanford's coreNLP is a collection of tools for processing human language (Manning et al., 2014). Specifically, I am using the constituency parser and its dependencies in order to obtain the syntactic analysis of the sentences input from the Hansard. The Stanford coreNLP parser

output is a collection of syntactic trees. Figure 2 shows an example of a syntactic tree from the output of the constituency parser. It utilizes the Penn Treebank tag set for parts of speech. A database was implemented as a .csv file with the columns in Figure 3. The χ^2 statistic was used to test the hypotheses.

```
<parse>(ROOT
(S
(NP (DT This))
(VP (VBZ is)
(NP
(NP (DT a) (NN petition))
(SBAR
(WHNP (WDT that))
(S
(VP (VBD was)
(VP (VBN signed)
(PP (IN by)
(NP (CD 10,000) (NNS people))))))))))
(. .)))
</parse>
```

Figure 2, CoreNLP output.

Is passive	ARG 1	Relative	verb	ARG 2	status1	status2	compared	animacy 1	animacy 2	animacy	TDP	tags
FALSE	This bill		signed	the Canadian	NEW	NEW	EQUAL	0	1	ARG 1 < A	FALSE	VBZ VBN
TRUE	a petition to that		signed	hundreds of p	NEW	NEW	EQUAL	0	1	ARG 1 < A	FALSE	VBZ VBN

Figure 3, Partial row from database.

Table 3 shows the individual tags that make up the passive voice. A list consisting of the tags that correspond to the passive voice where the preposition *by* is present were defined. This

combination of tags included the copula (*to be*), the past participle, and the preposition *by*. Table 4 contains examples of the tag sets and the text they match in the corpus, where *x* is the semantic theme, *y* is the semantic agent, and *attacked* is the past participle used in the passive construction.

Table 3, Part of Speech Tags (Marcus et al., 1999)

	Tag	Part of speech
1.	IN	Preposition
2.	VBD	Verb, past tense
3.	VBG	Verb, gerund or present participle
4.	VBN	Verb, past participle
5.	VBP	Verb, non-3rd person singular present
6.	VBZ	Verb, 3rd person singular present

Table 4, Passive Tags and Examples of Matching Text

	Tag set	Example of matched text
1.	['VBD', 'VBN', 'IN']	"x [VBD was/were [VBN attacked [IN by[y]]]]"
2.	['VBZ', 'VBN', 'IN']	"x [VBZ are [VBN attacked [IN by[y]]]]"
3.	['VBP', 'VBN', 'IN']	"x [VBP is [VBN attacked [IN by[y]]]]"
4.	['VBD', 'VBG', 'VBN', 'IN']	"x [VBD was/were [VBG being [VBN attacked [IN by[y]]]]]"
5.	['VBZ', 'VBG', 'VBN', 'IN']	"x [VBZ are [VBG being [VBN attacked [IN by[y]]]]]"
6.	['VBP', 'VBG', 'VBN', 'IN']	"x [VBP is [VBG being [VBN attacked [IN by[y]]]]]"
7.	['VBD', 'VBN', 'VBN', 'IN']	"x [VBD had [VBN been [VBN attacked [IN by[y]]]]]"
8.	['VBZ', 'VBN', 'VBN', 'IN']	"x [VBD have [VBN been [VBN attacked [IN by[y]]]]]"
9.	['VBP', 'VBN', 'VBN', 'IN']	"x [VBD has [VBN been [VBN attacked [IN by[y]]]]]"

In the case of the past- and present-progressive passive voice, the gerundive form of the verb was included (Table 4, items 4-6). In the case of the present-perfect and past-perfect passive voice (Table 4, items 7-9), the past tense of the copula was used.

However, while the passive structures are always made up of these specific tags in this order, there are often intervening phrases that are not relevant to the data being collected, such as adverbial phrases, which can occur as the left sibling of the highest VP and can interfere with detecting the passive voice if they are not accounted for. This set of tags was defined as seen in Table 5. Table 6 shows an example of text where a tag would be skipped.

Table 5, Skipped Tags (Marcus et al., 1999)

Tag	Part of speech
RB	Adverb
CC	Coordinating conjunction
DT	Determiner
JJR	Adjective, comparative

Table 6, Example of Skipped Text

Tag	Sample text
RB	The suspect [VBD was [RB quickly] [VBN arrested [IN by [the police]]]]]

The trees from the Stanford Parser were processed using NLTK's ParentedTree module (Bird et al., 2009). The tags for each lexical item were collected, skipping the exceptions, and checking the set for a combination of either three or four tags matching to those specified in the passive tag set.

At this time, its parallel French text was also selected. The French sentence was evaluated for preservation or replacement of the passive voice using a regular expression. Because the

Hansard is not aligned by punctuation, it was necessary to cycle through the sentences of the block of French text with input from the user to select the correct one.

Noun-Phrase Arguments

The two noun-phrase arguments before and after the preposition *by* were extracted. Argument 1 and argument 2 required slightly different handling.

For argument 1, the parent nodes were searched to find the highest verb phrase (VP) dominating the past participle (VBN). The parent node of the highest VP was ensured to be either (S) or (SBAR), filtering out phrases with matching tags dominated by another noun phrase (NP), which would indicate adjectival use of the passive (i.e. He was struck by a stolen car driven by the suspect). The left siblings of that VP were searched until a noun phrase was found, accounting for any intervening items such as adjectival phrases, as described above.

Figure 4 is an example of one of Stanford's parse trees printed from NLTK to show the syntactic structure. The immediate left sibling of the highest VP isn't the first noun-phrase argument needed for data, but rather the adverb that modifies it. We are concerned with the NP *people*, not the adverb *anywhere* that follows it.

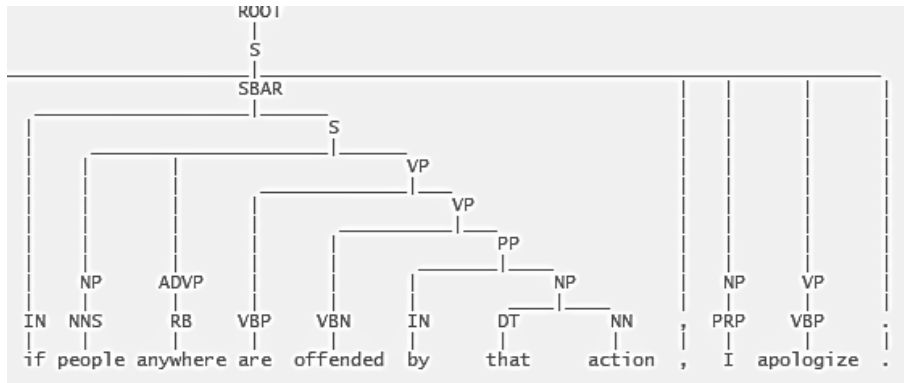


Figure 4, CoreNLP tree printed from NLTK.

Relative pronouns such as *that* and *which* were also accounted for at this stage. *Who* refers to a human entity and *which* refers to non-human entities. *That* may refer to either a human or non-human entity (Aygen, 2014). Given the definition of animacy used in this study – government organizations and industries which are considered animate and would be referred to with *which* or *that* – it is more useful to extract the noun phrase that the relative pronoun refers to in addition to the relative pronoun itself, rather than the relative pronoun alone, as the information it gives us about its referent is limited. Figure 5 shows the tree structure of a sentence containing a relative clause.

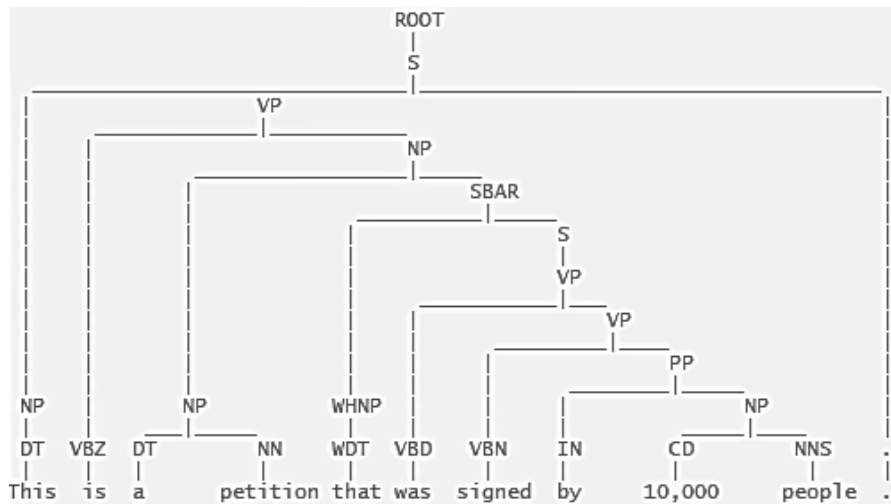


Figure 5, Relative clause.

For argument 2, the right sibling of the preposition *by* was searched. Siblings of *by* (IN) include (S), (SBAR) and (NP). However, only (S) and (NP) contain the required information; intervening SBARs, although a rare exception, had to be accounted for, such as in the sentence in Figure 6.

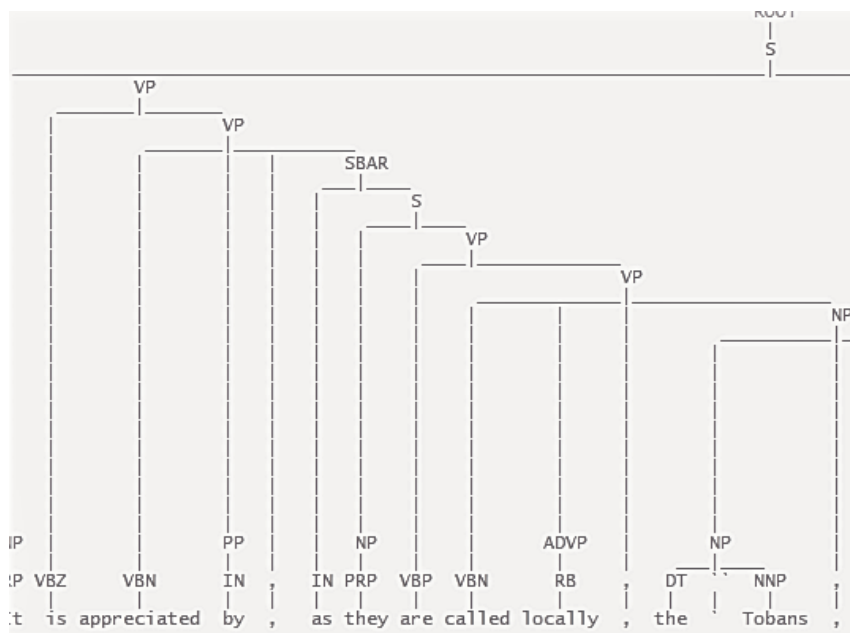


Figure 6, SBAR phrase.

The second argument of the *by* phrase, the noun phrase *the Tobans*, is not the immediate right sibling of *by* but is separated from it by the SBAR.

Information Status

Noun phrases were collected from the coreNLP parse trees and stored in a dictionary with a value of either *new* or *old*. Noun phrases were tracked for discourse status for the duration of the Hansard issue-- once a noun phrase had appeared once, it was considered discourse-old. This resolves several issues, including that pertaining to recurring entities that are inferrable and therefore considered discourse-old even when evoked in new contexts and noun phrases that represent different instances of the same item. It also has several limitations, including coreferential items and inferrable information that has not been explicitly evoked in the discourse.

Information status was only assessed according to whether the information is discourse-old or discourse-new. The topics of discussion in the Hansard are pre-determined and organized: “Discussion in the House of Commons must always be relevant to some definite proposal or motion” (ourcommons.ca). New noun phrases related to the topic of discussion coincide with the changing of speakers and topic. This means that anything pertaining to a topic of discussion raised in the Hansard could be considered hearer-old. It then makes sense to limit the current scope of this study to discourse status, or whether or not a particular noun phrase has previously been evoked in the discourse, rather than hearer status.

This doesn't mean it's impossible for a speaker to introduce an entity that is hearer-new. However, given that these communications are happening between a single speaker and many hearers, there is no way of knowing if the subject is new to all of them or what the threshold would be for considering something hearer-new if it is in fact known to some of them.

There are certain entities that recur regularly throughout the Hansard. Some examples are *the government, the chair, the party*, etc. Given that many of these frequently recurring entities can be inferred from the setting of the communication, the most useful way to categorize them seems to be to consider them inferrable information. As outlined in Prince's framework, this means they will be treated as discourse-old even if they haven't been explicitly evoked, or in this case, explicitly evoked in a new context. This category of entities can therefore be accurately kept track of using a dictionary.

There are also cases in which the same noun phrase may be referring to a different entity of the same class somewhere else in the discourse, for example, whether or not an entity such as *the bill* refers to a bill that has already been discussed or to a different bill that is being newly discussed. This made a possible case for resetting the dictionary upon every topic change; however, many entities can be carried across topics and speakers, which would erroneously mark as new many noun phrases that are actually old.

Again, recurring items such as bills or amendments can be considered inferrable given the setting in which the discourse is taking place. Additionally, in this particular corpus, the use of an entity to refer to another instance of its class is often signaled when pertinent. Bills are referred to by a specific alphanumeric code to differentiate them (*Bill S-14, Bill C-59*, etc.), and specific members of parliament (who are referred to as *my colleague, the member*, etc.) are

referred to by their riding (the location they represent) or other information signaling that a new entity is being referred to.

Limitations

Tracking the noun phrases according to when they first appeared was an effective way of assessing their discourse status in many situations. In the following example, both *Bill S-14* and *the other place*² are accurately marked as discourse-old, as they have been mentioned before.

Bill S-14 was adopted by *the other place* as tabled...
(Dechert, 2013)

However, issues arose with entities that were coreferential but represented by different sets of lexical items in the same span of text. In the following example, *the only transnational body of water* was correctly marked as appearing for the first time, but erroneously marked as discourse-new, as its referent, *Lake in the Woods*, appears earlier in the sentence:

(21) It is interesting that *Lake in the Woods_i* is perhaps the only *transnational body of water_i* that is not protected by the International Joint Commission. (Rafferty, 2010)

This matter was further complicated by entities that weren't explicitly evoked in the discourse but can be inferred from either world knowledge or the surrounding discourse. For example, in a discussion about the various safety measures that had been taken to lower the rate of car thefts:

(22) I believe that the studies at the time showed that if the *big car companies* were mandated by the government to install these immobilizers, it could have been done 10 years ago... (Maloway, 2010)

The program marks *big car companies* as discourse-new, as car companies have not been mentioned before in the discourse; however, it's safe to assume that from world knowledge and

² *The other place* is used by the House of Commons to indirectly refer to the Senate.

the surrounding discourse about cars the members of Parliament are already aware that such entities exist. More pragmatic information is needed from the surrounding context to accurately detect inferrable information and treat it accordingly as discourse-old. Additionally, it may be more useful to be able to classify information about these types of entities according to the more recent taxonomy proposed by Birner (2013). According to Birner, psycholinguistics research distinguishes between *forward* and *backward* inferences.

Forward inferences are also called *elaborating* inferences since they elaborate on information already mentioned in the discourse. A *forward* inference licenses inferences to material that may be mentioned later. It is possible for an entity to be related to prior discourse without having been explicitly evoked. This means that it is possible for an entity to be discourse-old and hearer-new, contrary to my previous statements about this category.

In (21), the preceding discussion regarding safety measures involving cars licenses gives rise to the discourse-old and hearer-new entity *big car companies*. This also explains why *big car companies* can be preposed in a passive construction, even though it hasn't been explicitly evoked in the discourse. In *backward* inferences, the inference isn't made until the utterance of the inferrable, and the hearer must make a connection to prior discourse to understand what it connects to.

In (22), *the only transnational body of water* counts as discourse-old and hearer-old, but because the set of words used to represent the entity differs, a connection must be made by the hearer from *the only transnational body of water* to its referent, *Lake in the Woods*. Table 7 outlines this taxonomy. Descriptions of these categories are as follows:

- Brand new: not inferentially linked and not known to the hearer
- Unused: not inferentially linked, but known to the hearer

- Evoked: inferentially linked, but known to the hearer
- Bridging inferrable: inferentially linked, but not known to the hearer

Table 7, Discourse Status (Based on Birner 2013).

	Discourse-new	Discourse-old
Hearer-new	Brand-new	Unused
Hearer-old	Bridging Inferrable	Evoked: Identity/ Elaborating Inferrable

Animacy

For this study, it was most useful to categorize noun phrases as a binary, given a value as either *animate* or *inanimate* based on user input. These values were then compared to assess whether the first argument of the *by* phrase was more animate than the second one, the second argument of the *by* phrase was more animate than the first, or if they were equal.

Tour de Présentation

The very qualities of the Hansard that make it an interesting corpus for this study are also those that exemplify the challenges that are posed by natural language. Manually inspecting each sentence in French for the location of noun phrases in order to compare it to its English parallel

is time consuming. Use of a translation API was necessary in order to make a more efficient and accurate attempt to locate the noun phrases in the parallel French sentences automatically.

Several options were considered in order to increase the likelihood that a *tour de présentation* could be found with the aid of a translation API in a corpus that is not a word-for-word translation. Methods used by translators that are necessary for conveying a clear meaning, such as *étouffement* and *effacement*, add another level of complexity.

One option was to translate the full English noun phrase into French and search for it in the parallel French sentence. Given the complexity of some of the noun phrases, and the freedom that translators have with conveying them, the exact set of lexical items that represents an entity is often not present in the parallel text, although the entity itself may be. This would have provided too narrow a possibility for locating the parallel noun phrase in French.

Another option was to translate the head of the English noun phrase. However, the improvements to machine translation allow it to benefit greatly from context, and translating a single noun removed from its context includes the high possibility of an inaccurate translation. Without context, the following types of errors can occur: *the chair* being translated as *la chaise* rather than *le président* in the appropriate situation.

Currently, the most reasonable approach appears to be to collect the nouns making up the noun phrase and search each one in the full French sentence translated into English via the API. This allows for the translation of the nouns in context but also provides a greater chance that one of them will be found in the parallel text despite minor variances in word choice. Nouns were chosen for the search as they contain the information more essential to the discourse than surrounding adjectives, and the verb is prone to undergo one of the several options to replace it in order to avoid or preserve the passive voice. While there may be cases where adjectives to

provide useful semantic information, such as *the Liberal Party* vs. *the Conservative Party*, it was not necessary for this data. In future work, the search will be expanded to include these features.

In translation of the passive voice, noun phrases may be kept in place, flipped using a *tour de présentation*, or one of them may be deleted. All four of these possibilities had to be accurately detected.

First, from the two extracted noun-phrase arguments, all of the nouns were collected. For example, *The chair of the Port of Montreal* would return a list of nouns as follows: [“chair”, “port”, “Montreal”].

The machine-translated French-to-English sentence was returned as a list words. The first item from the list of nouns was searched in the list of translated words. If it wasn't found, the next noun on the list would be searched, and so on. The positions of the nouns found in the machine-translated sentence were returned. The positions were compared. If the position of the first argument was greater than the position of the second argument, then a *tour de présentation* had occurred. If not, the noun phrases had remained in place. If none of them were found, then the noun phrase was deleted.

Limitations

Relying on machine translation in a study about choices made by human translators raised several issues. One was that although the entity represented by the noun phrase in the English sentence was present in the French sentence, it could be represented by a different set of lexical items than what the machine translation produced. This means that the program would

falsely flag the structure as having had one of its arguments deleted, or having no *tour de présentation*.

Some of the situations were easily dealt with, such as the variance in use of certain characters:

[E] agro-tourism -> [F] agrotourisme

A similar problem that arose from orthographic representation was the use of acronyms for organizations in place of the full name:

[E] DFAIT -> [F] le ministère des Affaires étrangères
the Department of Foreign Affairs

However, it was the use of synonyms that could make the parallel noun phrase difficult to detect.

There were also situations in which an entity was represented by a different but still semantically related set of lexical items that weren't produced by the machine translation. These went undetected in the machine-translated French-to-English sentence as a result. A few such examples are shown in Table 8.

Table 8, Related Words

English	French	French to English translation (machine)
Actions	Acte	Act
Vintners	Vinicoles	Wineries
Amendment	Changement	Change

RESULTS

The animacy of the noun-phrase arguments of the *by* phrase was shown to be a significant factor in determining whether or not a *tour de présentation* was used to translate the passive sentence ($\chi^2 = 12.74$, $p = .0123$), as shown in Table 9.

Table 9, Animacy of Arguments Compared, $\chi^2 = 12.74$ $p = .012$

	Arg1<Arg2 No RC	Arg1<Arg2 RC	Equal No RC	Equal RC	Arg1>Arg2
TDP	17 (15%)	1 (1%)	7 (6%)	0 (0%)	2 (2%)
Original order	24 (20%)	17 (15%)	24 (20%)	7 (6%)	17 (15%)

Table 10 shows the passive verbs that were replaced in the English-to-French translation using a *tour de présentation*.

Table 10, Passive Verbs in English Replaced by Using a *Tour de Présentation*

ignored (twice) experienced appreciated baffled envisioned heard encouraged	introduced covered laid (groundwork) undertaken taken	passed (information) mentioned informed made (a decision)	pursued approached violated reappointed handled	disputed owned established prorogued finalized
---	---	---	---	--

Most of the instances of the first argument being more animate than the second have a noun phrase in the Experiencer role as the first argument and a Theme noun phrase as the second argument.

I know [a number of my liberal colleagues [EXPERIENCER]]
were greatly disturbed by [the betrayal of the promise
[THEME]]...

(Nathan Cullen, Canada Elections act)

In the two occurrences of a *tour de présentation* being used to place a less animate noun phrase in front of a more animate one, the Experiencer was a first-person pronoun:

E: [**I** [EXPERIENCER] am baffled by *his position* [THEME].]

F: [*l'explication du directeur général des élections par
intérim* [THEME]...] [**me** [EXPERIENCER] sidère.]

[*the acting Chief Electoral Officer's explanation* [THEME]...] [**me** [EXPERIENCER].] (May, 2018)

However, all other occurrences of Experiencer/Theme roles in passives with *by* phrases maintained their original word order in translation, at 15% compared to 41% of those where the second argument was more animate.

Relative clauses were not completely exempt from the effects of syntactic shifts that are seen in other uses of the passive voice in *by* phrases. However, as predicted, their word order was much more tightly constrained. Whether or not the *by* phrase was contained in a relative clause was shown to be significant ($\chi^2 = 8.34$, $p = .003$), as shown in Table 11. The single example of a *tour de présentation* being used to translate a *by* phrase in a relative clause was used to move an animate noun phrase to the front of the sentence.

Table 11, *By Phrases in Relative Clauses vs Regular Clauses*
 $\chi^2 = 8.34$, $p = .003$

	No Relative Clause	Relative clause
TDP	24 (25%)	1 (1%)
Original order	48 (50%)	24(25%)

Other

Most of the cases fit neatly into two categories: a *by* phrase with two noun-phrase arguments that either changed places around the verb or remained in the same order. There were several examples that did not fit into these two categories. These included examples where one, or even both, of the arguments were deleted, or some alteration was made to the semantic characteristics of one of the arguments.

There were only a few cases where arguments were deleted, rather than the noun phrases being maintained or changed. There were not enough examples of this to make generalizations about the circumstances under which the deletion of an argument occurs. In the example below, the more animate second argument is deleted, rather than moving it to the front of the sentence:

E: **Bill C-4** was established by *the government* to kill two pieces of legislation...

F: **Ce projet de loi** visait à abroger deux mesures législatives...

This bill was intended to repeal two pieces of legislation... (Gérard Deltell, 255, Federal labor relations act)

Several examples of alterations to the characteristics of the noun phrases were found, including the following example, where *Statistics Canada*, which being an organization would be considered animate, is replaced by *service from Statistics Canada* in the French translation:

We are well served by Statistics Canada
 Nous bénéficions d'un bon service de Statistique Canada
 We benefit from good service from Statistics Canada 77.1
 (Murphy, 2010)

In some cases an animate noun phrase was fronted, but the changes that occurred to make this happen can't be classified neatly as either two noun phrases staying in place or changing position:

E: [I have heard the representations from **the hon. member**,]
 [as **similar representations** were made by *the hon. member*
before him.]

F : [**le député** a fait **des remarques similaires** à celle du
député qui a pris la parole avant lui.]

[**The hon. Member** made **similar remarks** to those of *the*
member who took the floor before him.] (Goodale, 2010)

Similar remarks isn't deleted, but rather attached to *the hon. member* rather than *the hon. member before him*.

The pronoun *on* is often used to replace passives with expletive *it* subjects, e.g., *it is often repeated / on l'a répété souvent*. In the following example it is used in combination with an apparent *tour de présentation* to place the first-person pronoun at the end of the sentence and the inanimate noun phrase *its support from across the aisles* towards the front of the sentence. However, the indefinite pronoun, which represents an animate and human subject, is inserted at the front of the sentence and the first-person pronoun placed within a relative clause.

E: [The last petition,] [and **I** am encouraged by *its support from across the aisles...*]

F: [La dernière petition,] [*pour laquelle on manifeste de l'autre côté de l'allée un appui qui **m'**encourage...*]

[The last petition,] [*for which ones shows from the other side of the aisle a support which encourages **me**...*]
(May, 2013)

More data is needed about these types of cases to see what factors determine when and where they are applied.

SCOPE

For this study I looked at written text in the form of speeches published in French and English, investigating factors that cause a syntactic shift to occur in translation. The specific difference I investigated was the translation of *by* phrases using a *tour de présentation*. The factors studied included noun-phrase animacy, information status, (specifically discourse-old and discourse-new), and whether or not the *by* phrase was contained in a relative clause.

FUTURE WORK

Understanding the conditions under which syntactic structure may be affected has the potential to benefit machine translation as well as sentence generation. This work is important to natural language generation because a proposition may be expressed by several sentences, and it is important to understand what factors determine the most effective way to say something, as well as what sentence structure is most suitable to particular situations.

The problems with tracking discourse status and detecting the use of a *tour de présentation* are related to information that comes from surrounding discourse. I would like to be able to measure the relevance of the current noun phrase to the discourse by organizing the noun phrases into a multi-parented tree structure, such as the one available in NLTK. This would allow for the identification of semantic hierarchies within text rather than assessing information in a linear fashion to determine discourse status.

The passive voice may be used to place emphasis on the entity undergoing the action described in a sentence. Authors claim that an entity is given emphasis in French by placing it at the end of the sentence using a *tour de présentation*. Understanding the surrounding pragmatic information may help to determine what degree of emphasis has been placed on an entity and how that factors in to determining a syntactic shift in the translation.

REFERENCES

- Aygen, Gülşat. *English Grammar: A Descriptive Linguistic Approach*. Kendall Hunt, 2014.
- Bally, Charles. "Intonation et syntaxe." *Cahiers Ferdinand de Saussure*, no. 1, 1941, pp. 33-43.
- Bird, Steven, Loper, Edward and Klein, Ewan (2009),
Natural Language Processing with Python. Available at <http://www.nltk.org/book>.
- Birner, Betty J. *Introduction to Pragmatics*. Wiley-Blackwell, 2013.
- Blinkenberg, Andreas. *L'Ordre des mots en français moderne*. 3rd ed, Munksgaard, 1928.
- Darbelnet, Jean and Jean-Paul Vinay. *Stylistique comparée du français et de l'anglais*. Didier, 1958.
- Delisle, Jean. *La traduction raisonnée*. University of Ottawa Press, 2003.
- Fillmore, Charles. *The Case for Case*. In Bach, Emmon, Harms, *Universals in Linguistic Theory*. Rinehart and Winston, New York, 1968.
- Grellet, Françoise. *Initiation à la version anglaise*. Hachette Livre, 1985.
- Levin, Beth. *English Verb Classes and Alterations*. University of Chicago Press, 1993.
- Manning, Christopher D., Surdeanu Mihai, Bauer, John, Finkel, Jenny, Bethard, Steven J., and McClosky, David. The Stanford CoreNLP Natural Language Processing Toolkit In: *Proceedings of 52nd 2014 Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55-60
- Marcus, Mitchell, et al. Treebank-3 LDC99T42. Philadelphia: Linguistic Data Consortium, 1999.
http://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html.
- Radford, Andrew. *An Introduction to English Sentence Structure*. Cambridge University Press, 2009.

Ellen Prince. *The ZPG letter: Subjects, definiteness, and information-status*, 1992. In William C. Mann and Sandra A. Thompson (Eds.), *Discourse Description: Diverse Linguistic Analyses of a Fund-raising text*. John Benjamins, Amsterdam, pages 295-325.
<https://doi.org/10.1075/pbns.16.12pri>.

Speeches

- Cullen, Nathan. “Canada Elections Act”. Hansard vol. 148, no. 255, 1st session, 42nd Parliament, Parliament of Canada, House of Commons. 2018.
Available at www.ourcommons.ca/DocumentViewer/en/42-1/house/sitting-255/hansard.
- Davies, Don. “Federal Public Sector Labour Relations Act”. Hansard vol. 148, no. 255, 1st session, 42nd Parliament, Parliament of Canada, House of Commons. 2018.
Available at www.ourcommons.ca/DocumentViewer/en/42-1/house/sitting-255/hansard.
- Dechert, Bob. “Fighting Foreign Corruption Act”. Hansard vol. 146, no. 272, 1st session, 41st Parliament, Parliament of Canada, House of Commons. 2013.
Available at www.ourcommons.ca/DocumentViewer/en/41-1/house/sitting-272/hansard.
- Deltell, Gérard. “Federal Public Sector Labour Relations Act”. Hansard vol. 148, no. 255, 1st session, 42nd Parliament, Parliament of Canada, House of Commons. 2018.
Available at www.ourcommons.ca/DocumentViewer/en/42-1/house/sitting-255/hansard.
- Gerretsen, Mark. “Federal Public Sector Labour Relations Act”. Hansard vol. 148, no. 255, 1st session, 42nd Parliament, Parliament of Canada, House of Commons. 2018.
Available at www.ourcommons.ca/DocumentViewer/en/42-1/house/sitting-255/hansard.
- Goguen, Robert. “Response to the Supreme Court of Canada Decision in R. v. Tse Act”. Hansard vol. 146, no. 224, 1st session, Parliament of Canada, House of Commons. 2013.
Available at www.ourcommons.ca/DocumentViewer/en/42-1/house/sitting-224/hansard.
- Goodale. “Federal Public Sector Labour Relations Act”. Hansard vol. 148, no. 255, 1st session, 42nd Parliament, Parliament of Canada, House of Commons. 2018.
Available at www.ourcommons.ca/DocumentViewer/en/42-1/house/sitting-255/hansard.
- Maloway, Jim. “Tackling Auto Theft and Property Crime Act Act”. Hansard vol. 146, no. 77, 3rd session, 40th Parliament, Parliament of Canada, House of Commons. 2010.
Available at www.ourcommons.ca/DocumentViewer/en/40-3/house/sitting-77/hansard.
- May, Elizabeth. “Canada Elections Act”. Hansard vol. 148, no. 255, 1st session, 42nd Parliament, Parliament of Canada, House of Commons. 2018.
Available at www.ourcommons.ca/DocumentViewer/en/42-1/house/sitting-255/hansard.
- May, Elizabeth. “Fighting Foreign Corruption Act”. Hansard vol. 146, no. 272, 1st session, 41st Parliament, Parliament of Canada, House of Commons. 2013.
Available at www.ourcommons.ca/DocumentViewer/en/41-1/house/sitting-272/hansard.

- Morneau, Bill. “Income Tax Act”. Hansard vol. 148, no. 012, 1st session, Parliament of Canada, House of Commons. 2016.
Available at www.ourcommons.ca/DocumentViewer/en/42-1/house/sitting-12/hansard.
- Murphy, Brian. “Serious Time for the Most Serious Crime Act”. Hansard vol. 146, no. 77, 3rd session, 40th Parliament, Parliament of Canada, House of Commons. 2010.
Available at www.ourcommons.ca/DocumentViewer/en/40-3/house/sitting-77/hansard
- Rafferty, Jim. “Lake in the Woods and Rainy River Basin”. Hansard vol. 146, no. 77, 3rd session, 40th Parliament, Parliament of Canada, House of Commons. 2010.
Available at www.ourcommons.ca/DocumentViewer/en/40-3/house/sitting-77/hansard.
- Sgro, Judy. “Iraq”. Hansard vol. 146, no. 062, 1st session, 40th Parliament, Parliament of Canada, House of Commons. 2011.
Available at www.ourcommons.ca/DocumentViewer/en/41-1/house/sitting-62/hansard.
- Trudeau, Pierre. “Statement on the introduction of the Official Languages Bill, October 17, 1968” *Library and Archives Canada*, Government of Canada, 2001, Available at www.collectionscanada.gc.ca/primeministers/h4-4066-e.html.