How Animacy and Information Status Determine Word Order in Translation of the Passive Voice

Ashli Fain, Reva Freedman

Department of Computer Science
Northern Illinois University
DeKalb, IL
ashlifain@gmail.com, rfreedman@niu.edu

Abstract

English uses the passive voice more frequently than French. One method of translating the passive includes rendering the sentence as active by using an active verb, and changing the placement of the verb's arguments. We are studying extra-syntactic features that predict where this method of translating the passive voice is used, including animacy and information status. We have obtained data from examining the Hansard, the transactions of the Canadian Parliament, which is published in both languages. This paper presents the results of a small mechanized corpus analysis on the relevance of the relative animacy of the agent (or experiencer) and the theme. This information will help to achieve desired stylistic output in a bilingual surface realizer.

Introduction

It is generally agreed that English uses the passive voice more frequently than French. Most textbooks on translation (e.g., Vinay and Darbelnet 1958, Delisle 2003) give syntactic alternatives that can be used in French to replace the English passive. We are studying extra-syntactic features that predict how an English passive will be rendered in French, for a future project involving a bilingual surface realizer in a dialogue system. We need to know which extra-syntactic features are relevant to generation choices so that we can keep track of those features during dialogue generation.

In this paper we examine the theoretical effect of differences in animacy and information status. We obtained data from examining the Hansard, the transactions of the Canadian Parliament, in which every parliamentary speech is published in both languages. In this paper we present the results of a small mechanized corpus analysis on the effect of relative animacy on the translation of the passive.

The passive voice poses a challenge for translators from English to French because there are many options for translating the passive and we want to maintain the desired stylistic qualities of the passive voice. We would like to be able to account for the difference in options so that we can choose the desired tone in our dialogue system.

Section 2 explains the theory behind our analysis. Section 3 describes the methodology of our study. Section 4 contains the results. Section 5 contains a discussion of the results, and Section 6 contains some conclusions and suggestions for future work.

Background and Related Work

The factors that are important for determining where the passive voice can be used and how it is translated include syntactic constraints, the relative animacy of the constituents, and their information status.

In French, the passive voice is used less frequently than in English (Delisle 2003). At the syntactic level, only accusative constructions can be passivized in French (Vinay and Darbelnet 1958). On the other hand, French has many alternatives outside of the passive voice to achieve the impersonal tone used in professional and scientific writing. According to Jisa (2002), the choice of passive or active voice cannot be considered simply a matter of grammatical difference, but is a matter of the speaker/writer's stylistic choices.

Experts give several methods of avoiding the use of the passive in French. One option is to maintain word order and choose an appropriate active verb, such as in (1), where the passive *was created* has been replaced by a different verb.

	Discourse-old	Discourse-new
Hearer-old	**Evoked: Identity/Elaborating** Inferrable (inferentially linked, and known to the hearer)	**Bridging Inferrable** (inferentially linked, but not known to the hearer)
Hearer-new	**Unused** (not inferentially linked, but known to the hearer)	**Brand new** (not inferentially linked and not known to the hearer)

Table 1: Discourse Status (Birner, 2013)

(1) E: **The delay** was created by *a financial downturn…*
F: **Ce retard** est dû à *un revers financier…*
(**this delay** is due to *a financial downturn…*)

Another option is to choose an active verb that requires a change in word order. Swapping the noun phrases in this fashion is called a *tour de présentation*. Formally, the case frame of the new verb requires arguments in a different order than the original passive. The relevant semantic roles in this study are the following: agent (the deliberate instigator of the action denoted by the verb), experiencer (the animate entity undergoing the effects of the action), and theme, which undergoes the action but does not change state. In a *tour de présentation* used to remove a passive verb, the agent or experiencer moves from the *by*-phrase to the syntactic subject, while the theme, which was the syntactic subject, becomes the syntactic object, as in (2).

(2) E: **Bill C-55** was studied by *the Standing Committee on Justice and Human Rights...*
F: Le Comité permanent de la justice et des droits de la personne a étudié **le projet de loi C-55**...
(*The Standing Committee on Justice and Human Rights* has studied **Bill C-55**...) (Goguen, 2013)

Additional options for translation of the passive voice into French include replacement of the verb by a reflexive form (e.g., *l'idée se montre*, lit. "the idea shows itself", for *the idea is shown*) or use of the indefinite pronoun *on* ("one", e.g., *on voit* for *it is seen*). Further structural changes can be added on top of these basic categories. Taking into account the factors that determine where these changes are applied is required to generate sentences that fit in the context of a continuous discourse or an ongoing dialogue.

In addition to syntactic criteria, we examined two non-syntactic criteria. Animacy of noun phrases is an important category in both syntactic and morphological analysis of natural language (Zaenen 2004). The animacy of a noun phrase determines which semantic roles it can fill (Fillmore 1968).

The animacy of the noun phrases along with the verb type have an effect on the choice of a speaker to passivize a sentence in English (Ferreira 1994). For verbs that can accommodate both an animate and inanimate syntactic subject, speakers prefer to place an animate NP in the subject position. The determination of the subject thus determines whether the sentence will be passive or not. If different rules apply in French and English, then the respective surface realizers need to use different rules to generate sentences with equivalent intent.

Information status has also been cited as a determining factor in word order (Blinkenberg 1928). Blinkenberg divides constituents into known/old and unknown/new information. He claims that French has a strong preference for placing known information first, and that this can be a motivation for translators to reconfigure the word order of a sentence in translation. Birner (2013, ch. 8) states that a constituent that is preposed in a passive construction must be equal to or older than the postposed constituent. Thus Blinkenberg's rule alone cannot be sufficient: if the original sentence were in the passive, the syntactic subject must already be old, and Blinkenberg's rule would never be applied.

Birner (2013, ch. 8) extended Blinkenberg's ideas by adding inferentially-linked constituents to Prince's given/new framework (Prince 1988). In Birner's approach, shown in Table 1, the given/new feature applies separately to the discourse and to the hearer's mental model. Discourse-old or new refers to whether the item has been mentioned before in the discourse, while hearer-old or new refers to whether the idea is new to the hearer. Any item that can be inferred from a previous discourse referent is considered discourse-old. In this way a constituent can be hearer-new while being discourse-old if it can be inferred from the discourse but has not actually been mentioned. Such a "bridging inferrable" satisfies the unknown/new part of Blinkenberg's criterion, while still fulfilling Birner's constraint that the preposed constituent of a passive must be as old as or older than the preposed constituent.

We are interested in information status because we would like to know whether certain combinations of discourse status force a change in word order or just give translators greater freedom with word order. If the latter is the case, then there is more freedom to express stylistic differences through word order without changing meaning.

Relative animacy:	Arg1 < Arg2	Equal	Arg1 > Arg2
Tour de présentation	26 (20%)	10 (7%)	2 (2%)
Original order	39 (30%)	27 (21%)	25 (19%)

Table 2: Word Order Based on Relative Animacy ($\chi^2 = 10.235$, p = .006)

Most generation systems have not tried to implement style-based distinctions. For example, SimpleNLG is a well-known surface realizer for English (Gatt and Reiter 2009). Vaudry and LaPalme (2013) modified SimpleNLG to generate both French and English, but choosing which constructions to use in French for the English passive was not part of their system.

Methodology

We performed a small experiment to see whether the above theoretical considerations about animacy were borne out in a corpus.

In order to identify how the English passive voice was translated into French, we selected from the Hansard, from March 3, 2010 to March 17, 2010, all the passive sentences originally in English that contained two noun phrases, one for agent or experiencer and one for theme. We used a Python program that searched for the following sequence: a noun phrase, a copula, a verb with a past participle suffix, the preposition *by*, and a second noun phrase, with any interpolated words. We verified by hand that the desired sentences were included.

Using the same sections of the Hansard, we also manually identified passive sentences that did not meet the above criterion, either because the agent was not specified (no *by*-phrase) or because no theme was specified, e.g., because the sentence contained expletive *it* as subject.

For the sentences that contained both an explicit agent and theme, these noun phrases in the English sentence were then categorized according to animacy. For the purpose of this study, every NP was categorized as animate or inanimate. In addition to individual politicians and members of organizations, the government and its branches were also classified as animate. In other words, any object whose parts were animate was considered animate via metonymy. Inanimate NPs in the corpus included bills, amendments, ideas and actions.

The French sentence was compared to the order of the NPs in the English sentence. We identified whether the agent and theme NPs occurred in the same order in the French sentence as in the English one.

We did not examine information status in this experiment because our software cannot yet identify the components of discourse status automatically.

Results

Table 2 shows the results for the 129 passive sentences in our corpus that contained two NPs. In 65 sentences (50%), the first argument was more animate than the second argument. Of these 65 sentences, the French translator changed the order to put the animate NP first in 26 of them (40%). In 39 sentences (60%), the word order and syntax remained unchanged.

Of the 27 sentences where the second argument was less animate than the first, i.e., the syntactic subject was already more animate, only 2 translations (7%) contained a changed word order. When the two noun phrases were of equal animacy, an intermediate percent of translations (10 sentences, or 27%) contained a changed word order. Out of the 38 sentences translated using a *tour de présentation*, 26 of them (68%) contained a second NP argument that was more animate than the first.

Thus the sentences most likely to be translated using a *tour de présentation* contained an inanimate NP as the first argument and an animate NP as the second. Sentences containing NPs of equal animacy were less commonly translated using a *tour de présentation*, and sentences where the second NP was less animate than the first were translated in this way only rarely. Although animacy alone may not be enough to motivate a *tour de présentation*, absence of an animate NP does seem to rule it out.

We used the χ^2 test to evaluate these results. The null hypothesis was that the relative animacy of agent and theme did not affect the percent of sentences where word order or syntactic form changed. The null hypothesis was rejected with $p < .01$, showing that the relative animacy of agent and theme is a highly significant factor in determining whether a change in word order and/or the use of the passive will occur.

Discussion

In addition to the statistical significance of animacy shown above, French seems to have a preference to front explicitly evoked (discourse-old and hearer-old) animate constituents wherever possible. The trend appears to be that constituents of similar information status are more like to be translated using a *tour de présentation*, whether evoked, unused, or inferred.

Given that a little under half of the noun phrases do not flip where the second argument is more animate than the first, animacy alone cannot account for the use of a *tour de présentation*.

However, on occasion the qualities of the noun phrases are altered in order to make them equal in animacy. In (3), although *des adeptes de l'intolérance religieuse*, being animate, would make an acceptable candidate for a *tour de présentation*, that option is not chosen. Instead, *the safety of 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 they were undergoing some action. This change elevates the theme (*les chrétiens*) to the level of animacy of the agent. This indicates that instead of moving the already animate agent to the front of the sentence, the theme may be made more animate and then moved to the front of the sentence.

(3) E: ...**the safety of Christians** was yet again threatened by *those pursuing a path of religious intolerance*.

F: ...les chrétiens ont de nouveau été menacés par *des adeptes de l'intolérance religieuse*.

(...**Christians** were threatened yet again by *those pursuing a path of religious intolerance*.) (Sgro, 2011)

While the preference in French to place the agent of an action first is strong, passive sentences with themes and agents are frequently translated as such into French. If the preference to place an agent at the front of the sentence were so strong as to always effect a *tour de présentation*, the expected effect would be that all passive sentences in English would be translated in French by placing the thematic agent first. However, this is not the case. Thus there is still room for stylistic choice, or possibly for other factors we have not identified.

Conclusions and Future Work

In this paper we have shown that when agent and theme are both represented in a passive English sentence, the relative animacy of these two items is a highly significant factor in determining how the sentence will be translated into French. This translation can occur via a *tour de présentation*, where the syntactic subject and object switch places, or via an idiosyncratic transformation.

In future work, we would like to see whether the same patterns in varieties of French other than Canadian French. Since it is possible for members of the Canadian parliament to give speeches in either English or French, we would also like to look at what happens to passive French sentences when translated into English. We would also like to study *tours de présentation* used for reasons other than eliminating a passive. Finally, we would like to mechanize

the determination of information status so that it can be included in the corpus analysis. All of these would help us identify cases where word order must change and where it is optional. The former will help us generate idiomatic sentences in both languages; the latter will help us identify stylistic variation and its causes.

References

Birner, B. J. 1988. *Introduction to Pragmatics*. Hoboken, NJ: Wiley-Blackwell.

Blinkenberg, A. 1928. *L'ordre des mots en français moderne*, 3rd ed. Copenhagen: Munksgaard.

Delisle, J. 2003. *La traduction raisonée*. Ottawa: University of Ottawa Press.

Ferreira, F. 1994. Choice of Passive Voice is Affected by Verb Type and Animacy. *Journal of Memory and Language* 33(6): 715-736.

Gatt, A., and Reiter, E. 2009. SimpleNLG: A Realisation Engine for Practical Applications. In *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG '09)*, 90-93. Stroudsburg, PA: ACL. http://www.aclweb.org/anthology/W09-0613.

Goguen, R. 2013. Response to the Supreme Court of Canada Decision in R. v. Tse Act. Hansard vol. 146, no. 224, 1st session, 41st Parliament, Parliament of Canada, House of Commons. www.ourcommons.ca/DocumentViewer/en/42-1/house/sitting-224/hansard.

Jisa, H., Reilly, J., Verhoeven, L., Baruch, E., and Rosado, E. 2002. Passive Voice Constructions in Written Texts: A Cross-Linguistic Developmental Study. *Written Language and Literacy* 5(2): 163-181. https://doi.org/10.1075/wll.5.2.03jis.

Prince, E. 1992. The ZPG Letter: Subjects, Definiteness, and Information-status. In W. C. Mann and S. A. Thompson (eds.), *Discourse Description: Diverse Linguistic Analyses of a Fund-raising Text*, 295-325. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.16.12pri.

Sgro, J. 2011. Iraq. Hansard vol. 146, no. 062, 1st session, 40th Parliament, Parliament of Canada, House of Commons. www.ourcommons.ca/DocumentViewer/en/41-1/house/sitting-62/hansard.

Vinay, J.-P., and Darbelnet, J. 1958. *Stylistique comparée du français et de l'anglais*. Paris: Didier.

Vaudry, P.-L., and LaPalme, G. 2013. Adapting SimpleNLG for Bilingual English-French Realisation. In *Proceedings of the 14th European Workshop on Natural Language Generation*, 183-187. Stroudsburg, PA: ACL. http://www.aclweb.org/anthology/W13-2125.

Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitina, T., O'Connor, M. C., and Wasow, T. 2004. Animacy Encoding in English: Why and How. In *Proceedings of the ACL '04 Workshop on Discourse Annotation*. Stroudsburg, PA: ACL. http://www.aclweb.org/anthology/W04-0216.