(15) Our professor is *on* the bus.

As discussed earlier, this sentence is ambiguous in that the professor could be riding the bus, or he could be located on the roof of the bus. This sentence is translated into two French sentences, one for each case: [6]

(16) Notre professeur est *dans* le bus.

(17) Notre professeur est *sur* le bus.

In (16), the professor is riding the bus, while in (17), he is located on the roof of the bus. This type of ambiguity is detected in the same way as simple conceptual ambiguity, the only difference being that in the complex case, all the French conceptual representations do not have the same descriptive parts.

6 Conclusion

In this paper, we have described a system of translation for locative prepositions that uses Herskovits' idea of the ideal meaning of prepositions and Lakoff's idea of ICM's. While our work does not prove the linguistic and psychological theories on which it is based, it suggests that they can be useful in machine translation. We chose to use conceptual knowledge to deal with the translation of locative prepositions, first, because it provides an elegant solution to the problem, and second, because we believe that conceptual knowledge of the sort that we use could be useful in other cognitive tasks such as story understanding, vision, and robot planning.

7 Acknowledgments

We wish to thank Graeme Hirst for invaluable comments and detailed readings of many versions of this work, and to gratefully acknowledge the financial support of the Department of Computer Science, University of Toronto, and the Natural Sciences and Engineering Research Council of Canada.

8 References

[Grimaud 1988] M. Grimaud, "Toponyms, Prepositions, and Cognitive Maps in English and French," *Journal of the American Society of Geolinguistics*, vol. 14, pp. 54-76, 1988.

[Herskovits 1986] A. Herskovits, *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English*, Cambridge University Press, Cambridge, MA, 1986.

[Japkowicz 1990] N. Japkowicz, "The Translation of Basic Topological Prepositions from English into French," M.S. Thesis, published as *Technical Report CSRI-243*, University of Toronto, 1990.

[Japkowicz & Wiebe 1990] N. Japkowicz & J. Wiebe, "Using Conceptual Information to Translate Locative Prepositions from English into French," *Current Trends in SNePS—Proceedings of the 1990 workshop,* Ali, Chalupsky, Kumar (eds.), forthcoming.

[Lakoff & Johnson 1980] G. Lakoff & M. Johnson, *Metaphors we Live by*, University of Chicago Press, Chicago, 1980.

[Lakoff 1987] G. Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, University of Chicago Press, Chicago, 1987.

[Rosch 1977] E. Rosch, "Human Categorization," in *Advances in Cross-Cultural Psychology*, vol. 1, N. Warren (ed.), pp. 1-49, Academic Press, London, 1977.

[Zelinsky-Wibbelt 1990] C. Zelinsky-Wibbelt, "The Semantic Representation of Spatial Configurations: a conceptual motivation for generation in Machine Translation," *Proceedings of the 13th International Conference on Computational Linguistics,* vol. 3, pp. 299-303, 1990.

[6] In sections 1, 2, and 3, only the first case was considered.

Of the 35 cases, only 3 are in the specific category. Of the remaining, 18 cases are in the semi-general category and 14 are in the general category.

5 Error and Ambiguity Detection

The conceptual representations that were designed for the purpose of translation can also be used to detect certain kinds of errors and ambiguities. Below, we describe two kinds that can be detected by the system: conceptual errors and conceptual ambiguity.

5.1 Conceptual errors

The system can detect two types of conceptual errors: conceptualization errors and usage errors or abnormalities.

5.1.1 Conceptualization errors

Conceptualization errors occur when the preposition requires the reference object to be conceptualized in a way that it cannot be in the language considered. An example of a sentence where such an error occurs is

(11) * The boy is at the shade.

This sentence is erroneous because 'at' requires 'shade' to be conceptualized as a point, but 'shade' used as a reference object can never be conceptualized as a point in English. This error can be detected by the system because no conceptual representation of shade as a reference object is built whose conceptualization is point. This error is detected in the first step of the second phase of the system.

5.1.2 Usage errors and abnormalities

Usage errors and abnormalities occur when the demands of the preposition are satisfied by the reference object, but the conditions required of the located object by the conceptual representation, or general conditions required of all types of relations, are not. Such an error occurs in the following:

(12) * The man is in the board.

The use of 'in' is fine, considering just the reference object; for example, a nail can be located *in* a board. The problem is that the located object is 'man', and a man cannot be embedded in a board under normal circumstances. This error is detected by the system because the condition on the located object (in the conditional part of the conceptual representation) is not verified. This error is detected in the second step of the second phase of the system.

5.2 Conceptual ambiguities

Conceptual ambiguity is ambiguity where the English preposition has several meanings in French. The system can detect two types of conceptual ambiguities: simple and complex. Both are detected during the first step of the second phase of the system.

5.2.1 Simple conceptual ambiguity

In the case of simple conceptual ambiguity, an ambiguous English preposition is translated into a single French preposition that is ambiguous in the same way. For example:

(13) The boy is *at* the supermarket.

Sentence (13) can be understood to mean either that the boy is shopping at the supermarket, or that he is on a trajectory going by the supermarket, and is currently located at the supermarket. Its French translation is

(14) Le garçon est *au* supermarché,

which carries the same ambiguity as the English sentence. This type of ambiguity is detected when several English conceptual representations can be instantiated for a single sentence. All instantiated English conceptual representations have identical descriptive parts. In the case of simple conceptual ambiguity, all the French conceptual representations happen to have the same descriptive part.

5.2.2 Complex conceptual ambiguity

The difference between simple and complex conceptual ambiguity is the following: in the former, the French sentence carries the same ambiguity as the English sentence, but in the latter, the ambiguity is not carried through the translation (so the English sentence has two different French translations). Complex conceptual ambiguity is present in (6), which is repeated here as sentence (15):

4.2.2 Derivation of the objective meaning of the sentence

This phase is also performed in two steps. The first step identifies the English conceptual representations relevant to the sentence, according to the preposition used. That is, given the ideal meaning of the preposition used in the English sentence, certain conceptual representations that were built in the previous phase are discarded. In example (6), the only conceptual representation of a bus that will remain is that of a surface, since the ideal meaning of 'on' allows the reference object to be a surface or a line and, while a bus is sometimes conceptualized as a surface, it is never conceptualized as a line.

The second step discards even more conceptual representations, this time based on the type and/or properties of the located object. In sentence (6), no conceptual representation is discarded at this point. This is so because the only condition on the located object is that it can be supported by the reference object, and this condition is verified for (6) because a human being can be supported by a bus. In sentence (7), however, the conceptual representations of a bus as a surface are discarded because an elephant cannot be supported by a bus.

The second step also builds the objective meaning of the sentence. The objective meaning of a sentence is derived from the conceptual representation chosen in the first step of this phase. Its main component is the *properties* field. This *properties* field has the same type of content as the *properties* field of the conceptual representations. It is this shared field that allows a matching between the English conceptual representation and an objective meaning.

In certain cases, in this step, several objective meanings can be derived. In these cases, the sentence is ambiguous (see section 5).

4.2.3 Derivation of the target-language preposition

This phase has, once again, two steps. The first consists of matching the objective meaning of the sentence to a French conceptualization. This can be done in a way similar to that of the previous step: by matching the *properties* field of the objective meaning of the sentence with the *properties* field of the French conceptual representation of the reference object.

The second step consists of matching a French preposition to the French conceptual representation derived by the previous step. This is done in a straight-forward way, using a look-up table. In example (6), the French conceptualization is matched to the preposition 'dans'.

4.2.4 Finalization

The Finalization phase consists of only one step: that of generating the French sentence. In example (6), it is at this point that the French version, "Notre professeur est *dans* le bus", is generated.[5]

4.3 Coverage

We implemented the system on a large number of cases, where each case is an "objective situation", such as an object being on a horizontal support or an object being in a closed environment. There are 35 cases, which can be divided into the following three categories:

- *Specific*, i.e., cases in which the reference object is a given object; the expressions 'on the wall' (meaning against the wall), 'at sea', and 'in the air' are the specific cases in the system.

- *Semi-general*, i.e., cases in which the reference object belongs to a well defined category of objects. Examples are being in a country (e.g., 'in England' and 'in France') and being in a piece of clothing (e.g., 'in a hat', 'in a shirt', and 'in a pair of shorts').

- *General*, i.e., cases in which the reference object belongs to an abstract category of objects. Examples are being on a planar surface (e.g., 'on the table', 'on the floor', 'on the chair', and 'on the roof') and being at an artifact with a given purpose (e.g., 'at the door', 'at his books', 'at his desk', and 'at his typewriter').

[5] Note that we are not taking ambiguity into consideration here. If we were, then the sentence "Notre professeur est *sur* le bus." would also be generated (meaning that our professor is on the roof of the bus). This case will be discussed in section 5.

point of focus. At its current state, the system deals solely with isolated sentences, so it is unable to perform this checking.

3.2 The descriptive part

The descriptive part of a conceptual representation includes the following three types of information about the conceptualization: its *dimension*, its *fullness*, and its *width*.

* Its *dimension* is the main information about the conceptualization. The possible values of the dimension field include *point*, *line*, *surface*, and *volume*.

* Its *fullness* can take the values *empty* or *full*. Fullness is important when, for example, the dimension is *volume*. Consider the following sentences.

(9) The girl is in the tree.

(10) The nail is in the tree.

One needs to differentiate between the situation of (9), in which the located object (the girl) is located in the tree, and the one of (10), in which the located object (the nail) is embedded in the tree. This distinction, however, is not needed to translate between English and French (it might be needed with other languages, though); rather, it is needed to understand the sentence.

* Its *width* takes the values *existent* or *non-existent*.[4] Width is important for sentences such as those in (4), where the width is conceptualized as being non-existent in French, and existent in English, this difference leading to a difference in the use of the locative prepositions (French uses 'sur' and English uses 'in').

[4]Remember that the descriptive part describes conceptualizations. Therefore, when we describe the width to be existent or non-existent, it is the width in the conceptualization that is in question, not that of the real object. Objectively, for example, a picture has a width, but this width is so small that it is ignored in some of its conceptualizations. Objectively also, a picture is the re-creator of an environment. The conceptualizations in which this objective property is highlighted have an existent width, since environments can contain 3-dimensional entities.

4 The Algorithm

4.1 Overview

Our method of translation first transforms the source-language sentence into a source-language representation (the English conceptual level), and then translates the source-language representation into a target-language representation (the French conceptual level). This target-language representation is finally used to generate the target-language sentence. The algorithm works in four phases:

1. Initialization
2. Derivation of the objective meaning of the sentence
3. Derivation of the target-language preposition
4. Finalization

4.2 Phases

In the description that follows, each step is explained and illustrated with example (6).

4.2.1 Initialization

The initialization phase is composed of two steps. The first consists of parsing the input sentence and returning some information about each noun, such as its role in the sentence (located or reference object), its French translation, and certain useful French morphological and syntactic information about it. In sentence (6), for example, this information is that 'Our professor' is the located object, that its French translation is 'Notre professeur', and that 'professeur' is a masculine common noun in French; and also that 'bus' is the reference object, that its French translation is 'bus', and that 'bus' is a masculine common noun in French.

The second step consists of building the conceptual representations of the located and reference objects (see Japkowicz 1990 and Japkowicz & Wiebe 1990). All possible conceptual representations are built at this point—the discrimination of those that are relevant to the sentence from the others is done in the next phase.

about the conditions necessary for the conceptualization to hold.

A conceptual representation of an object is composed of a conditional part and a descriptive part. The conditional part is a list of properties of the object and of its situation in the sentence. The former kind of property is objective information about the object, such as its shape, the parts it is made of, and its function. The latter properties are whether the object is a located or reference object, and whether the sentence is in English or French. The descriptive part is a description of a conceptualization of that object. This part is conceptual, rather than objective. Here follows a detailed description of conceptual representations.[2]

3.1 The conditional part

The conditional part is made up of the following types of properties:

* The *role* in the sentence of the object being considered (located or reference object). [3]

* The *language* in which the sentence is uttered (English or French). This condition is crucial to the system because not all conceptualizations are possible in both languages, and these differences account for differences in use of the prepositions. This point is important, for example, for pairs of sentences (4), where a picture is conceptualized as a volume in English and as a surface in French; for pairs of sentences (5), where the shade is conceptualized as a volume in English and as a point in French; and for pairs of sentences (6), where a bus is conceptualized as a surface in English and as a volume in French.

* The *properties* of the reference object that are relevant to the objective spatial relation expressed in the sentence (these properties are language independent). This part of the conceptual representation specifies the objective situation in which the object being conceptualized is involved. It is central to the system because it is common to English and French (since it describes an objective situation) and is the part of the conceptual representation that allows a matching between English and French. For example, consider (4). The properties of a picture that are relevant given the objective meaning of the sentence are the fact that it is the re-creator of an environment, with entities included in that environment, and that it is an object with a very small, almost non-existent, width. These properties are common to English and French. What differs are the conceptualizations: English highlights the first property, conceptualizing the picture as a volume, while French highlights the second, considering the width to be non-existent and conceptualizing the picture as a surface.

* *World-knowledge conditions* involving the located object of the sentence (for example, whether the located object can be supported by the reference object). These conditions are used to check the plausibility of a sentence with respect to the located object. For example, the sentences in (6) are plausible, while the sentence

(7) The elephant is on the bus

is not, since an elephant is too heavy to be supported by a bus. In general, this condition is used to check for abnormalities within one language rather than to account for differences between English and French. Section 5 describes how the system detects such abnormalities.

* *Extra-sentential constraints.* Extra-sentential constraints are pragmatic constraints, derived from the context in which the sentence is uttered, that can influence the choice of preposition. For example:

(8) The gas station is *at* the freeway. [Herskovits 1986, p. 138]

This sentence is valid only when the speaker pictures himself or herself as being on a trajectory intersecting the reference object at the

[2] Certain aspects of the conceptual representations were implemented for extensibility or for the purposes of ambiguity and error detection. For the sake of completeness, we describe all aspects in this section, even those not directly related to translation (see Japkowicz 1990 for further explanation of these aspects).

[3] Note that a located object is *always* conceptualized as a point. This is so because the conceptualization of the located object has no impact on the use of the prepositions. It is the conceptualization of the reference object that is relevant.

appropriate for expressing a relation between a point and a surface, and French uses 'dans' (the French equivalent of 'in'), the preposition appropriate for expressing a relation between a point and a volume. The appropriateness of a preposition for expressing a certain relation is determined by its ideal meanings.

2.2.3 Our synthesis

Our task consisted of synthesizing Herskovits' and Grimaud's contributions and making this synthesis suitable for a computational system, since both Herskovits and Grimaud's analyses are mainly linguistic and not directly geared towards computation.

Our first task was to define the ideal meanings of each preposition:

AT/À:

- relation between two points.

ON/SUR:

- relation between a point and a surface whose boundaries are irrelevant.
- relation between a point and a line.

IN/DANS:

- relation between a point and a bounded surface.
- relation between a point and an empty volume.
- relation between a point and a full volume.[1]

Our next task was to develop a knowledge representation of a conceptualization of an object, that is, a representation of the way an object can be conceptualized, given a particular language, a particular situation, etc. Typically, in our application, these conceptualizations are geometrical objects, such as points, lines, surfaces, and volumes.

[1] Note that Herskovits' notion of ideal meaning involves more information than ours: rather than the vague term 'relation', Herskovits identifies the specific sort of relation that holds between the two objects, such as coincidence, support, and containment. For the specific problem in translation that we address, such specifications are unnecessary. They would be necessary, however, in a system designed for a deeper understanding than ours is designed to achieve.

Our final task was to design a system of translation. Our system works as follows: given the source-language sentence, its *objective meaning* (i.e., its language-independent meaning) is derived. This is done by first using the ideal meanings of the source-language preposition to find the conceptualization that applies to the reference object, and then deriving the objective meaning of the sentence from this conceptualization. (Because each conceptualization of an object used as a reference object corresponds to some objective meaning, this last step is easily performed.) Given the objective meaning of the sentence, the conceptualization of the reference object that should be used in the target language is then found. Finally, using the list of ideal meanings of the target-language prepositions together with the target-language conceptualization, the system derives the preposition to be used in the target-language sentence.

2.2.4 Other work

Independently, Zelinsky-Wibbelt (1990) took an approach similar to ours to the problem of translating locative prepositions. She worked on translation between English and German rather than English and French. This supports our hypothesis that the theory we use can be extended to pairs of languages other than English and French.

In addition to the types of expressions our system translates, her system translates sentences with verbs other than 'to be'. The reason why we chose not to process sentences using verbs other than 'to be' was to study the prepositions themselves in detail, before addressing the more complicated problem of their interactions with verbs. Zelinsky-Wibbelt does not refer to any preliminary detailed study of the prepositions themselves. We carried on a detailed bilingual study of locative prepositions by adapting and expanding the case studies of Herskovits (1986).

3 The Conceptual Representation of Objects

The central entity in our research is the *conceptual representation of objects* (or *conceptual representation*), which represents a conceptualization together with information

⟨located object⟩ ⟨*be*⟩ ⟨locative preposition⟩ ⟨reference object⟩

⟨located object⟩ ⟨locative preposition⟩ ⟨reference object⟩

2.1 Examples of the problem

While in the most representative uses of locative prepositions, there is a direct correspondence between English and French ('in' corresponding to 'dans', 'on' to 'sur', and 'at' to 'à'), in many cases, this correspondence does not hold.

The following pairs of sentences illustrate cases in which the correspondences hold:

(1) The boy is *in* his room.
Le garçon est *dans* sa chambre.

(2) The glass is *on* the table.
Le verre est *sur* la table.

(3) The secretary is *at* her desk.
La secrétaire est *à* son bureau.

Sentences (4), (5), and (6), in contrast, illustrate cases in which the correspondences do not hold:

(4) My friend is *in* the picture.
Mon ami(e) est *sur* la photo.

(5) The lounge chair is *in* the shade.
La chaise longue est *à* l'ombre.

(6) Our professor is *on* the bus.
Notre professeur est *dans* le bus.

At first sight, the correspondence between English and French locative prepositions may seem arbitrary. Our analysis, however, reveals that coherence might be found.

2.2 Analysis of the problem

Our analysis takes its principal sources in the works of Herskovits (1986) and Grimaud (1988).

2.2.1 Herskovits' contribution

Herskovits (1986) contributed to the solution to our problem by introducing the concept of the *ideal meaning* of a locative preposition. This concept is inspired by Rosch's (1977) prototype theory, in which human categorization of objects is viewed as organized around prototypes (best instances of the category) and distances from these prototypes (the shorter the distance of an object away from a prototype, the more representative of the category the object is). In the case of prepositions, *prototypical* or *ideal meanings* are geometrical relations between the *located object*, the object whose location is being specified in the sentence, and the *reference object*, the object indicating the location of the located object.

A second contribution of Herskovits is her case study of the three locative prepositions 'in', 'on', and 'at'. Our own study of 35 different cases is heavily based on this part of Herskovits' work.

2.2.2 Grimaud's contribution

Grimaud (1988) presents a linguistic analysis of locative prepositions in English versus French. His theory is based on Lakoff & Johnson (1980) and Lakoff (1987) and uses the notion of *conceptualizations* of objects. A conceptualization is a mental representation of an object or an idea which takes into consideration not only the "objective truth" about that object or idea, but also human biological perception and experience.

In his theory, Grimaud suggests that the cases in which the correspondences described in section 2.1 do not hold are not simply exceptional but rather are due to differences in the ways that English and French conceptualize the objects involved in the relation. The reason why the same object can be conceptualized as different geometrical objects in different languages, given a particular situation, is that objects have several properties (or aspects) and different languages might not choose to highlight and hide the same properties (or aspects) of a given object in a given situation. This happens in (6), for example (under the interpretation in which the professor is riding the bus rather than being located on the roof of the bus)— English conceptualizes the bus as a *surface* that can *support* entities, by highlighting only its bottom platform, while French conceptualizes the bus as a *volume* that can *contain* entities, by highlighting its bottom surface, its sides, and its roof altogether. This leads to a difference in the way that English and French express the spatial relation: English uses 'on', the preposition

A SYSTEM FOR TRANSLATING LOCATIVE PREPOSITIONS FROM ENGLISH INTO FRENCH*

Nathalie Japkowicz
Department of Computer Science
Rutgers University
New Brunswick, NJ 08903
nat@yoko.rutgers.edu

Janyce M. Wiebe
Department of Computer Science
University of Toronto
Toronto, Canada M5S 1A4
wiebe@cs.toronto.edu

Abstract

Machine translation of locative prepositions is not straightforward, even between closely related languages. This paper discusses a system of translation of locative prepositions between English and French. The system is based on the premises that English and French do not always *conceptualize* objects in the same way, and that this accounts for the major differences in the ways that locative prepositions are used in these languages. This paper introduces knowledge representations of *conceptualizations* of objects, and a method for translating prepositions based on these conceptual representations.

1 Introduction

This paper presents an analysis of the differences in the uses of locative prepositions in two languages, and then describes an automatic system of translation that is based on this analysis.

Our research originated from the observation that even between two closely related languages such as English and French, locative prepositions of even simple sentences do not seem to be translated from one language to the other in a clearly systematic and coherent way. However, the translation becomes more coherent if we introduce Herskovits' idea of the *ideal meaning* of a preposition (Herskovits 1986) and Lakoff's idea of Idealized Cognitive Models (ICM's) (Lakoff 1987). A central part of our research was to design entities based on Lakoff's ICM's. We call these entities *conceptual representations of objects*. The main thesis of this paper is that, even though the ideal meanings of the locative prepositions we studied are the same in English and in French, these two languages do not always conceptualize the objects involved in a scene in the same way and that this leads to differences in the translation of locative prepositions. This theory seems suitable to pairs of languages other than English and French, as well.

In addition, we will also describe how the system detects abnormalities and ambiguities using knowledge required for the translation task.

This paper is organized as follows: section 2 presents an analysis of and a solution to the problem of translating locative prepositions from English into French, section 3 presents the conceptual representations of objects, section 4 presents the algorithm we designed and implemented for translating locative prepositions, section 5 discusses the detection of abnormalities and ambiguities, and section 6 is the conclusion.

2 Translating Locative Prepositions

We now describe the differences between English and French locative expressions and give a possible analysis of the problem. Specifically, we concentrate on the translation of the three locative prepositions 'in', 'on', and 'at', into the French prepositions 'dans', 'sur', and 'à', in the context of simple sentences or expressions of the form:

*The research described in this paper was conducted at the University of Toronto.