Contextual Resolution of Global Ambiguity for MT

Cathy Berthouzoz

University of Texas at Austin
Cathyb@cs.utexas.edu

University of Geneva
Berthouzoz@lettres.unige.ch

University of Texas at Austin
Department of Philosophy
Campus mail code C3500
Austin/TX 78712

Abstract

In this paper, we explore the integration of context into domain-independent machine translation based on a pseudo-semantic approach. We choose a theory of discourse structure (SDRT) to provide the model of context. The incompatibility between the knowledge-poor translation model with the knowledge-rich discourse theory leads to supplement the first with some basic lexical semantics, and to replace the specific rules of the second with more general pragmatic principles. How contextual information is used to choose the preferred interpretation of globally ambiguous sentences is then described with two examples.

Introduction

Language is a means of communication. As speaker/writer of a language we use our linguistic and world knowledge to choose words that best express what we mean – or do not mean. Words are then combined into bigger units to form sentences, which in turn are combined into texts. As hearer/reader, we use the same knowledge to infer what the speaker/writer intended to communicate. The task of translation involves yet one more step: expressing within the target language what was communicated in the source language.

It is a matter-of-fact that machine translation is not yet equipped with the necessary tools to adequately translate real texts in unrestricted domains. Such a task will become feasible once we will be able to come up with interpretation and disambiguation models that can cope with huge amounts of linguistic and extra-linguistic knowledge, while minimizing computational costs. However, a small step toward this idealistic, long-term goal is to switch from sentence translation to *sentence-in-context* translation.

The *sentence-in-context* translation approach follows a pseudo-semantic transfer model (Etchegoyen 1998; Etchegoyen and Wehrli 1998) in the sense that the functional words (prepositions, determiners, conjunctions), along with the aspectual information, are given a full semantic interpretation, while the open class words (nouns, adjectives and verbs) are left uninterpreted. However, it departs from the pseudo-semantic transfer approach in the sense that contextual information is integrated into both the analysis and transfer processes.

During source analysis, the need for contextual information arises from the necessity to resolve anaphors in order to achieve a full interpretation of the sentence, in the first place, to resolve global ambiguity in order to select the most appropriate interpretation in the given context, in second place. As a result, the selected interpretation is included into the current context. During transfer, contextual information is needed to disambiguate multiple bilingual choices. The use of contextual information leads to another augmentation of the pseudo-semantic approach, in the sense that the semantic representation of the sentences will include some semantic information for the open class words. The integration of the context into the translation process is shown in Figure 1 below.

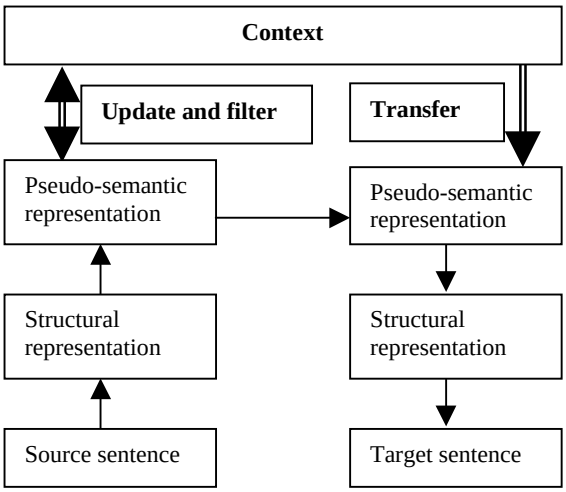

Figure 1 Model of *sentence-in-context* translation.
↑ = analysis, ↓ = generation

The context is heterogeneous – various types of information come into play during the interpretation of a sentence. Moreover, since it grows as the discourse goes on, a theory of dynamic discourse interpretation is chosen as framework, namely the *Segmented Discourse Representation Theory* (*SDRT*; Asher 1993). Thus, the

model of context is provided by the theory. However, the standard theory has to be adapted to the needs of domain-independent machine translation on the one hand, and to the specific model assumed here on the other. These adaptations will be described in the next section. Two examples of global disambiguation within this model will then be discussed.

Adaptation of SDRT to Machine Translation

SDRT constructs the discourse structure incrementally, by connecting each new sentence to the structure already built, by means of discourse relations. Although this theory is based on *Discourse Representation Theory* (*DRT*; Kamp and Reyle, 1993), it is better suited to machine translation, since the structure of the text is reflected in the representation of the discourse. It is not the case in DRT, where the whole discourse ends up with one huge flat structure. From the generation point of view, we need to have clues on the source text structure, because the target text should reflect the source structure as much as possible, within the target language adequacy (Doherty 1997).

Within SDRT, the semantic content of a clause is represented by a *Discourse Representation Structure* (*DRS*), which is a pair $<U_k, Con_k>$ where U_k is the set of discourse entities denoting individuals, eventualities and times, and Con_k is the set of conditions on those entities. The semantic content of a text, built up by connecting these DRSs with rhetorical relations, is represented by a recursive structure called a *Segmented DRS* (*SDRS*). An SDRS is a pair $<U_K, Con_K>$, where U_K is the set of DRS or SDRS and Con_K a set of conditions (discourse relations) on those (S)DRS. New material can be attached only to *open constituents*, i.e. constituents on the right frontier of the discourse (Asher 1993; Grosz and Sidner, 1986). Discourse relations are usually signaled by syntactic markers such as *as a result, then, however* (Knott 1995). In the absence of such cue phrases, discourse relations need to be inferred from the reader's knowledge base, especially world knowledge (Asher 1993; Lascarides and Asher 1993; Asher and Lascarides 1995).

Foreground and Background Contexts

When inferring a relation, the reader's knowledge base contains the SDRS of the text so far, the logical form of the current sentence, the assumption that the text is coherent, all the pragmatic and world knowledge, and all the laws of logic. Within standard SDRT, the knowledge base is thus unstructured. Intuitively, however, the contextual information can be partitioned into foreground and background information, the former being expressed in the discourse, the latter not[1]. Thus, the foreground context contains the SDRS of the text so far, while the background context is made up of the pragmatic and world knowledge[2]. If discourse relations are introduced by syntactic markers, it is not necessary to resort to the background context. The foreground context is domain-independent, and (partly) language-independent, while the background context is domain-dependent.

In order to infer the discourse relations, SDRT strongly relies on world knowledge. This is incompatible with a domain-independent translation system, since it would amount to implementing vast amounts of knowledge. Moreover, the translation model assumed here does not derive a semantic representation for the open class words. Hence, in order for SDRT to be meaningful within this model, a compromise has to be reached. Therefore, the pseudo-semantics will be supplemented with some lexical semantics necessary to infer the most important discourse relations. As a result, the background context will be kept as general as possible.

Discourse Relations

Another departure from standard SDRT is that not all the relations defined in the theory are inferable from the discourse. In the absence of syntactic markers that introduce discourse relations, the only relations that are inferred without resort to world knowledge are Narration, Background and Elaboration. Let $R(\alpha,\beta)$ express the fact that the clause β is connected to the clause α by relation R. Then the discourse relations and their entailments are defined by the following pragmatic rules, adapted from (Lascarides and Asher 1993), (Asher and Lascarides 1995), and (Asher et al. 1995):

- **Narration**: if β is to be attached to α in context τ, then, in the absence of any other information, Narration(α,β).
- **Background**: if β is to be attached to α in τ, and the eventuality of β is a state, then Background(α,β).
- **Elaboration**: if the eventuality of β is negated, then Elaboration(α,β).

The temporal implications that the discourse relations entail about the eventualities they connect are defined by the following rules:

- **Axiom on Narration**: if Narration(α,β), then the event of α precedes the event of β.
- **Axiom on Background**: if Background(α,β), then the state of β overlaps the event of α.

The causal implications that the discourse relations entail about the eventualities they connect are defined by the following rules:

- **Axiom on Explanation**: if Explanation(α,β), then the event of β caused the event of α.

[1] This distinction is also made in [illegible] contexts and knowledge base contexts.

[2] The assumption of coherence and the laws of logic are not part of the contextual information.

- **Axiom on Result**: if Result(α,β), then the event in α caused the event of β.

The following rules constrain the relation between Background and the other relations:

- **Constraint 1 on Background**: if Background(α,β), then β is closed for Narration.
- **Constraint 2 on Background**: if Background(α,β), then Background(γ,β) is not possible for $\alpha \neq \gamma$.

Narration being the default relation, incoherence arises when two clauses related by a Narration do not elaborate on a common topic. However, incoherence does not block the translation, since even an incoherent source text need to be translated into the target language. I will therefore assume a weak notion of Narration, in the sense that it does not trigger the creation of a topic constituent, even in the cases of coherent text.

I define a third rule that constrain the attachment of two background relations to the same constituent:

- **Constraint 3 on Background**: if Background(α,β), then Background(α,γ) is possible iff γ is consistent with β.

General Pragmatic Rules

Discourse relations associated with specific lexical rules are used to disambiguate word senses in (Asher and Lascarides 1995). In the pseudo-semantic approach word senses are expressed only in the bilingual lexicon. Thus, disambiguation will take place during the transfer phase. This is not a matter at issue here, although it is crucial to translation. On the contrary, global ambiguity has to be resolved before transfer, since it leads to more than one pseudo-semantic representation, and only one is transferred, the preferred one, if possible. To resolve this ambiguity with contextual information without relying on world knowledge is possible only with general pragmatic principles.

The first general principal states that the preferred interpretation is the one that is connected with the strongest discourse relation. The second principle states that a discourse relation is strengthened by anaphoric links. Anaphoric links can hold between a pronoun and its antecedent, or a referential expression and its antecedent. Referential expressions trigger presuppositions about the existence of their antecedent in the discourse universe. The third general principle states that there is inconsistency between two representations if they contain contradictory DRS conditions holding at the same time.

Basic Lexical Semantics

The above principles have lexical semantics consequences. First, the anaphoric link between a referential expression and its antecedent should hold even if the antecedent is implicit. As we will see in the second example below, the expression *the water* in the sentence "*I didn't stay long in the water*" refers to the water that is implicit to the activity of swimming in the previous sentence "*I went to swim*". The proposed solution is to supplement the lexical entry of the verbs of activity with the usual location of the activity. For example, swimming is normally done in water, flying is normally done in the air, and running is normally done on earth. This could be represented as a kind of shadow argument (Pustejovsky 1990), which would be accessible for presupposition binding. Second, if consistency is to be checked, the lexicon needs to include a relation for contradictories that hold between lexical items, such as cold/hot, big/small, etc…

Context at work

Natural language interpretation requires contextual information because ambiguity is one of the pervasive characteristics of natural language. Global ambiguity, for instance, arises when the analysis of a sentence generates more than one interpretation. Within a translation system, one interpretation has to be selected for further processing. Therefore, contextual information is used as a filter on the set of hypotheses. This role will be exemplified with two examples, an instance of literal/idiomatic interpretation, and an instance of anaphoric/expletive interpretation.

Example 1: He kicked the bucket

Even if the context is an essential part of natural language interpretation, it is not an inherent characteristic of natural language: words and expressions do actually have a meaning by themselves. We all know what the expression *He kicked the bucket* means outside any context. Our lexical knowledge tells us that *he* refers to a male individual that was previously introduced in the discourse, *kick* describes an action performed with one's foot toward a physical object, and that *the bucket* is a physical object, used as a container. Our grammatical knowledge tells us that the agent of the action is the male individual referred to by *he*, that the object on which the kicking was done is the specific object referred to by *the bucket*, and that this action was performed in the past and is actually finished. We can use this sentence to describe a specific man performing this physical action as expression of his anger, as in (1)below. But *kick* and *the bucket* together form an idiomatic expression meaning *to die*. Thus, we can use the same sentence to express the fact that a specific man has died, as in (2) below.

1. a) Mary came into the room with a bucket of water. b) John was angry. c) He kicked the bucket. d) As a result, the water spilled over the floor.
2. a) Max had a car accident last week. b) He was dead drunk. c) He kicked the bucket. d) As a result, his wife is staying alone with four kids.

When translating this sentence into French, the right interpretation has to be picked up in order to generate the correct target expression: the idiomatic expression "*Il a cassé sa pipe*" is not appropriate in (3) below, and the

literal expression "*Il a donné un coup de pied dans le baquet*" has nothing to do with the story in (4).

3. a) Marie entra dans la pièce avec un baquet d'eau. b) Jean était en colère. c) Il a donné un coup de pied dans le baquet./#Il a cassé sa pipe. d) En conséquence, l'eau s'est renversée sur le sol.

4. a) Max a eu un accident de voiture la semaine passée. b) Il était ivre mort. c) #Il a donné un coup de pied dans le baquet./ Il a cassé sa pipe. d) En conséquence, sa femme reste seule avec quatre gamins.

How to rule out the wrong interpretation? Let α, β, and δ be the representations of the sentences (1.a-b,d) above, and γ_l and γ_i be the literal and idiomatic representations of sentence (1.c), and τ the representation of the context, $\tau = \varnothing$ before processing (1.a), and $\tau = \{\alpha\}$ after. When processing (1.b), α is the only attachment point, thus we try to attach β to α, so we have the hypothesis $<\alpha,\alpha,\beta>$. The main eventuality of β being a state, we infer that β forms a background to α, i.e. the relation Background(α,β). This relation leads to the creation of a Foreground-Background Pair (FBP; Asher et al. 1995), combining the information contained in α and β, noted here $\alpha+\beta$. Background(α,β) entails that β temporally overlaps with α.

When processing (1.c), there are two possible attachment points, $\alpha+\beta$, and β, so we have the four hypotheses $<\tau,\alpha+\beta,\gamma_l>$, $<\tau,\alpha+\beta,\gamma_i>$, $<\tau,\beta,\gamma_l>$, $<\tau,\beta,\gamma_i>$. In the case of $<\tau,\alpha+\beta,\gamma_l>$, the only relation that can be inferred is Narration($\alpha+\beta,\gamma_l$), and the anaphor *he* can be resolved with *John*. The presupposition triggered by the referential expression *the bucket* can be bound within $\alpha+\beta$, thus strengthening the discourse relation Narration($\alpha+\beta,\gamma_l$). Narration(β,γ_l) or Background(β,γ_l) are ruled out by the Constraints 1 and 2 on Background. Result(β,γ_l) would hold if we had a law establishing the causality between being angry and kicking. This is a far too specific law to be included in the knowledge base, thus $<\tau,\beta,\gamma_l>$ is ruled out. In the case of the idiomatic interpretation, we infer Narration($\alpha+\beta,\gamma_i$). $<\tau,\beta,\gamma_i>$ is ruled out on the same grounds as for the literal interpretation. The preferred interpretation is thus γ_l, since it is connected to the context with the strongest relation.

The prediction is confirmed by sentence (1.d). Contrary to the previous sentences, the discourse relation is given by the syntactic marker *as a result*. As the only attachment point is γ, we infer Result(γ_l,δ) and Result(γ_i,δ) for the literal and idiomatic interpretation respectively. The referential expression *the water* triggers a presupposition that can be bound within γ_l through a referential chain linking *the bucket* in γ_l and *a bucket of water* in $\alpha+\beta$, in case of the literal expression, confirming the prediction. It cannot be bound in case of the idiomatic expression.

In case of discourse (2), the discourse structure is the same, but in this case, the presupposition triggered by the referential expression *the bucket* in the literal interpretation cannot be bound, thus the preferred interpretation is γ_i. The discourse structures of (1) and (2) are given in (5.a) and (5.b) respectively.

5. a) $K_{(1)}$:={{$\alpha,\beta,\gamma_l,\delta$}, {Background($\alpha,\beta$), Narration($\alpha+\beta,\gamma_l$), Result($\gamma_l,\delta$)}
   b) $K_{(2)}$:= {{$\alpha,\beta,\gamma_i,\delta$}, {Background($\alpha,\beta$), Narration($\alpha+\beta,\gamma_i$), Result($\gamma_i,\delta$)}

Example 2: It was too cold

Outside any context, the pronoun *it* is ambiguous between its expletive and anaphoric uses in the simple sentence (6.a). Therefore, the three sentences in (6.b) are possible translations into French. Adding context allows us to reduce the ambiguity. For instance, in discourse (7), intuitively, the preferred interpretation is that the water is too cold, although the weather is too cold is also possible.

6. a) It was too cold. b) Il faisait trop froid/ Il était trop froid/Elle était trop froide.

7. a) Yesterday, I went to swim. b) I didn't stay long in the water. c) It was too cold.

8. a) Hier j'ai été nager. b) Je ne suis pas restée longtemps dans l'eau. c) Elle était trop foide./Il faisait trop froid.

Let α, and β, be the representations of the sentences (7.a-b), γ_e and γ_a be the expletive, in this case the weather use, and anaphoric representations of sentence (7.c), and τ the representation of the context, as above. When processing (7.b), α is the only attachment point, thus we have the hypothesis $<\alpha,\alpha,\beta>$. From the negation in β, we infer Elaboration(α,β).

When processing (7.c), there is only one available attachment point, β, so we have the two hypotheses $<\tau,\beta,\gamma_e>$ and $<\tau,\beta,\gamma_a>$. The eventuality in γ being a state, we can infer Background(β,γ_e) and Background(β,γ_a). In the case of $<\tau,\beta,\gamma_e>$, the anaphor *it* can be resolved with *water*. The anaphoric link strengthens the relation, thus, the preferred interpretation is γ_a.

If no mention of water is made, as in discourse (9), there is no antecedent, the anaphor cannot be resolved, thus ruling out the anaphoric interpretation (10.c). The discourse structures of discourses (7) and (9) are given in (11.a) and (11.b) respectively.

9. a) Yesterday, I went to swim. b) I didn't stay long. c) It was too cold.

10. a) Hier, j'ai été nager. b) Je ne suis pas restée longtemps. c) Il faisait trop froid./#Elle était trop froide.

11. a) $K_{(7)}$:= {{α,β,γ_a}, {Elaboration(α,β), Background(β,γ_a)}
    b) $K_{(9)}$:= {{α,β,γ_e}, {Elaboration(α,β), Background(β,γ_e)}

If we change the context by adding some information about the weather, as in (12), the weather interpretation is ruled out. Let α, β, δ, γ_a and γ_e be the representations of the sentences of discourse (12). The main eventuality in α

being a state, we infer Background(β,α). There are two attachment points, due to the creation of a FBP, thus, we have the hypotheses <τ,α+β,δ> and <τ,β,δ>. For both, we infer Narration, thus we have Narration(α+β,δ), and Narration(β,δ). Although δ can be attached to two different points, it is the only open constituent for γ. In the case of <τ,δ,γ_a>, we can infer Background(δ,γ_a), and resolve the anaphor *it* with *water*. For <τ,δ,γ_e>, Background(δ,γ_e) is ruled out by the Constraint 3 on Background, and the principle of consistency, since the weather cannot be hot and cold at the same time. Thus, the preferred interpretation is γ_a (13.d).

12. a) Yesterday, it was really hot. b) I went to swim. c) I didn't stay long in the water. d) It was too cold.
13. a) Hier, il faisait très chaud. b) J'ai été nager. c) Je ne suis pas restée longtemps dans l'eau. d) #Il faisait trop froid./ Elle était trop froide.

Note that if the relation between δ and γ_e is Narration, as this is the case in (14) below, the weather interpretation is the preferred one (15.d). To treat discourse like this one requires a law stating that a big quantity of water cannot suddenly become cold. It is a too specific knowledge to be included in the system, which will thus make the wrong prediction. The discourse structures of (12) and (14) are given in (16.a) and (16.b) respectively.

14. a) Yesterday, it was really hot. b) I went to swim. c) I stayed in the water for a couple of hours. d) Then, it got cold.
15. a) Hier, il faisait très chaud. b) J'ai été nager. c) Je suis restée dans l'eau pendant plusieurs heures. d) Puis, il a fait froid. /#Puis, elle s'est refroidie.
16. a) $K_{(12)}$:= {{α,β,δ,γ_a},
{Background(β,α),Narration(α+β,δ),
Narration(β,δ), Background(δ,γ_a)}
b) #$K_{(14)}$:= {{α,β,γ_a},
{Background(β,α), Narration(α+β,δ),
Narration(β,δ), Background(β,γ_a)}

Concluding Remarks

In this paper, we have proposed to supplement a domain-independent translation system with the context provided by a theory of discourse interpretation, namely SDRT. However, the context as provided by SDRT is dependent on world knowledge, while the pseudo-semantic approach of the system does not provide the necessary semantics. We have proposed to supplement the lexicon with some lexical semantics, and to replace the specific rules of SDRT with more general pragmatic principles. Finally, we have shown how these principles work during global disambiguation, and pointed out the problem that with too little world knowledge, the model fails to make the right prediction in some cases. This is the starting point of further research about the refinement of lexical knowledge compatible with domain-independence.

Acknowledgments. The author gratefully acknowledges the Swiss National Science Foundation fellowship 81GE-54587. Many thanks go to Eric Wehrli, for steady encouragement, and to Ana Alves, Nicholas Asher, Tim Fernando, and Isabel Gomez Tzurruka for helpful discussions.

References

Asher, Nicholas. 1993. *Reference to Abstract Objects in Discourse*. Studies in Linguistics and Philosophy. Dordrecht: Kluwer Academic Publishers.

Asher, Nicholas; Michel Aurnague, Myriam Bras, Pierre Sablayrolles and Laure Vieu. 1995. De l'Espace-Temps dans l'Analyse du Discours. *Sémiotiques 9: Théories Sémantiques et modélisation*, 11-62. Didier-Érudition, Centre national de la recherche scientifique, France.

Asher, Nicholas and M. Morreau. 1991. Commonsense Entailment. *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, 387-392. Sydney, Australia. Los Altos, CA: Morgan Kaufman Press.

Asher, Nicholas and Alex Lascarides, 1995. Lexical Disambiguation in a Discourse Context. *Journal of Semantics* 12:69-108.

DYX%D 6D ªQLYORVH5 J Lexical Ambiguity Using a Formal Theory of Context. In Kees Van Deemter and Stanley Peters (eds), *Semantic Ambiguity and Underspecification*. CSLI Lecture Notes Number 55. Cambridge: Cambridge University Press.

Doherty, Monika. 1997. Textual Garden-Paths. In Christa Hauenschild and Susanne Heizmann (eds), *Machine Translation and Translation Theory*, 69-89. Berlin, NewYork: Mouton de Gruyter.

Etchegoyen, Thierry. 1998. Génération Automatique de Phrases. Le système GBGen. Technical Notes 98-1, Dept. of Linguistics, Geneva Univ.

Etchegoyen, Thierry, and Eric Wehrli. 1998. Traduction Automatique et Structures d'Interface. Proceedings of Traitement Automatique du Language Naturel, 2-11. Paris, France.

Grosz, Barbara J., and Candice. L. Sidner. 1986**.** Attention, Intentions, and the Structure of Discourse. *Computational Linguistics* 12(3): 175-204.

Kamp, Hans, and Uwe Reyle. 1993. *From Discourse to Logic*. Dordrecht: Kluwer Academic Press.

Knott, Alistair. 1995. A Data-Driven Methodology for Motivating a Set of Coherence Relations. Ph.D. diss., Dept. of Computer Science, Edinburgh Univ.

Lascarides Alex, and Nicholas Asher, 1993. Temporal Interpretation, Discourse Relations and Commonsense Entailment. *Linguistics and Philosophy* 16(5): 437-493.

Pustejovsky, James. 1998. *The Generative Lexicon*. Cambridge, Mass.: The MIT Press.