[From: *Konvens '94*: *Verarbeitung natürlicher Sprache*, Wien, 28-30 September 1994, hrsg. Harald Trost]

Towards Lexical Representation for Interlingua-based Machine Translation

Bianka Buschbeck-Wolf
Scientific Centre IBM Heidelberg
Vangerowstr. 18
69115 Heidelberg
e-mail: bianka@vnet.ibm.com

Abstract This paper addresses problems of lexical representation in Interlingua-based machine translation. Since lexical items in different languages do not share the same meanings, the definition of Interlingua concepts is not straightforward and in the case of generalization and specification gaps perhaps impossible. In current systems such conceptual divergences are usually treated as analysis or generation tasks. This results in the use of lexical representations which are not strictly monolingual. I will show how target language independency of lexical description can be preserved by making use of a module which negotiates between monolingual lexical representations. It detects the relationships between source and target language lexical items and determines the set of concepts shared by both of them. Thus, the Interlingua representation is derived with respect to a given contrastive situation. Those cases of lexical ambiguity that are preserved by the other language, generalization and specification gaps are recognized in advance so that the analysis is restricted to the necessary depth. These ideas are illustrated by the bilingual translation of spatial prepositions.

Zusammenfassung Dieser Beitrag beschäftigt sich mit Problemen der lexikalischen Repräsentation in der Interlingua-basierten Übersetzung. Da miteinander korrespondierende Lexeme verschiedener Sprachen nicht dieselben Bedeutungen umfassen, erweist sich die Definition von Interlingua-Konzepten als problematisch und im Fall von Generalisierungs- oder Spezifizierungslücken als nahezu unmöglich. Die übliche Verlagerung der Behandlung von Konzeptualisierungsdivergenzen auf Analyse bzw. Generierung führt dazu, daß die Zielsprachenunabhängigkeit der lexikalischen Beschreibung verlorengeht. Am Beispiel der bilingualen Übersetzung räumlicher Präpositionen wird ein Verfahren vorgestellt, mit dem die strikt monolinguale lexikalische Beschreibung erhalten werden kann. Vor der eigentlichen Analyse wird durch den Vergleich quell- und zielsprachlicher konzeptueller Repäsentationen die aktuelle Übersetzungssituation festgestellt und die Menge der interlingualen Konzepte identifiziert. Damit wird die Ebene des Überganges in die Zielsprache der kontrastiven Situation entsprechend determiniert. Sprachübergreifende Ambiguitäten, Generalisierungs- und Spezifizierungslücken werden erkannt, so daß die Analyse auf das notwendige Maß beschränkt bleibt.

1. Problems of the naive Interlingua approach to machine translation

The idea of the Interlingua (IL) approach is the mapping of a source language (SL) expression to a target language (TL) expression via a language independent representation which is shared by the languages involved. This method is especially promising for multi-lingual machine translation (MT) since the addition of new languages to the system does not require the addition of new transfer modules for all language pairs involved. This idea is also attractive for bilingual systems since it allows for bidirectionality. But "while the addition of new languages may appear easy in an interlingual system, there are major disadvantages: the difficulties of defining an interlingua, even for closely related languages." [HS92: 75] This problem concerns

especially the representation of lexical meaning in IL systems. It is caused by the fact that different languages conceptualize the world in different ways so that the meanings of words in these languages do not line up. Since lexical differences arise from conceptual ones the definition of completely language-neutral IL concepts is not straightforward. Only in the case when a SL and a TL lexical item share the same concept the idea of an IL is feasible. But there are also cases of so called "translation mismatches" [KOP91] or conceptual divergences, where one language lexicalizes more general or more specific concepts than the other one. Hence, translation becomes a matter of approximation. The use of a more general or more specific TL lexical item leads to the loss or the addition of information respectively.

In general, there are two ways to define interlingual concepts for lexical items. On the one hand, IL concepts are considered as specific as they are in the SL and the TL, i.e. the set of IL concepts equals the conjunction of all language specific concepts. On the other hand, IL concepts represent a general meaning which is subsumed by SL and TL lexical items without being specific to one of them. In the first case the set of IL concepts becomes very large and the mapping of a SL expression to the appropriate IL representation requires a great amount of analysis. Here, the identification of the TL lexical specification has to be provided already by the SL analysis. This also affects the meaning representation of the SL lexical items. It mirrors the conceptualization of the TL so that the TL independence assumption of the IL method is at risk. In the second approach the language specific conceptual differences are neutralized. Then, the generation is not straightforward if the TL makes a lexical distinction which is covered by one general IL concept. Here, the generator has to include techniques that assign preferences to a certain way of expressing a concept, see [BMR91]. In both approaches the problem of defining IL concepts in the case of "translation mismatches" remains unsolved. It is viewed either as an analysis or a generation task. This leads [KGN91: 78] to the conclusion: "there is no hope of constructing a language IL which is such that translations in all languages map onto a single abstract representation in IL". Even if the aims of the IL approach are confined to the design of a conceptual representation mediating between two languages there are many translation mappings which cannot share the same representation.

2. A modification of the naive Interlingua model

If we want to maintain the language-pair independence assumption for the lexical representation on the one hand and the module accounting argument on the other, the naive IL model has to be modified. In this section I show how this can be done.

Figure 1 presents a modified IL model where a special negotiation module for the identification of the contrastive situation is introduced. It has the following tasks:

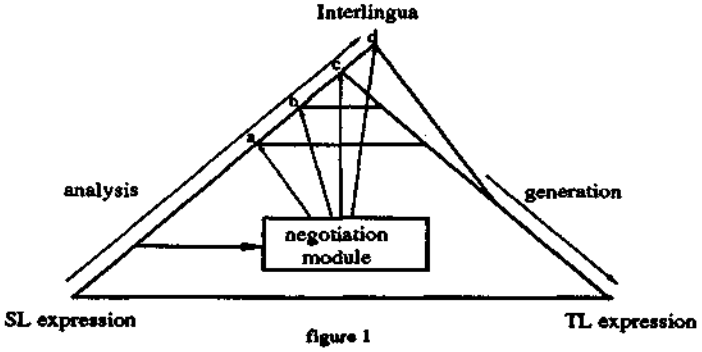

figure 1

1. Identification of possible TL correspondences for a given SL lexical item
2. Establishment of the relation between SL and TL lexical items
3. Derivation of an efficent analysis strategy

1. Before starting the analysis of an ambiguous lexical item its meaning representations are sent to the negotiation module. The negotiator looks for conceptual representations in the TL lexicon which overlap with each of the given SL meanings in order to pick out the possible TL correspondences. Their identification is based on identity, subset and superset relations between SL and TL conceptual meanings.
2. In the next step the relationship between the SL word and the found corresponding TL words is established. If there is only one TL lexical item which shares all the meanings of the SL lexical item, they are assumed to be *completely equivalent*. If some or each of the meanings of an ambiguous SL lexical item is shared by different TL words a *partial equivalence* is identified. While equivalence relations presuppose the identity of SL and TL meanings, *translation mismatches* are recognized via their non-identity. By making use of superset and subset relations between TL and SL meanings it can be inferred that the conceptualization in the TL is either more general or more specific than in the SL.
3. In the last step the analysis strategy is planned. The main ideas of this procedure can be summarized as follows. If SL and TL lexical items are identical in all their meanings, i.e. a *complete equivalence* relation is recognized, no analysis is required since the SL ambiguity is preserved in the TL. This is exemplified under case a in figure 1. The IL representation which is handed over to the generator equals the set of shared concepts. In the case of *partial equivalence* the analysis has to pick out the appropriate interpretation which is shared by the SL and the TL lexical item in the given context. This is shown in case c of figure 1. In the case of *generalization*, which is inferred from the superset relation between a TL and a SL meaning, the translation mapping is straightforward since the former implies the latter. This case is exemplified under b in figure 1. In the case of *specification* the SL conceptual representation is split into two or more specific representations in the TL. An analysis according to the TL specification is indispensible. The TL predicts the specifity of the IL representation and thus the depth of the analysis, cf. case d of figure 1.

By making use of the knowledge about the contrastive situation in the lexical domain, the analysis is guided by TL requirements and thus restricted to the necessary depth. The definition of an interlingual conceptual representation in the case of conceptual divergences is avoided by a mapping based on a *flexible Interlingua* which fits the actual situation (cf. figure 1). Furthermore, the important principle of the language-pair independence in the lexicon which, as shown in section one, cannot be realized in the naive IL model, is now preserved. From the use of strictly monolingual representations we gain the advantage that the same representation is applicable to analysis as well as to generation. However, the application of the proposed approach imposes a high degree of formalization on the conceptual representation in order to determine the relation between SL and TL expressions.

I will illustrate the approach outlined in this section with the German-English bidirectional translation of prepositions. Given the limited space I restrict their readings to the spatial domain. Assuming that the spatial as well as other interpretations are identified for the most part by sortal constraints on the arguments of the preposition (for detailed discussion see [BEH94]), I pick out a subset of spatial readings of some German and English prepositions which call for considerably deep representations and provide an interesting variety of translation phenomena. After having

introduced a formal representation language for spatial relations, I focus on issues that concern the monolingual representation of spatial prepositions in the lexicon. I present the translation problems we are faced with and show how they are dealt with in the proposed IL model.

3. Semantics of spatial prepositions

Let us consider spatial relations between two objects expressed by the German topological prepositions "in" ('in'), "auf" ('on'), "an" ('on', 'next to') and their English correspondences "in" and "on" respectively, as they occur in expressions of the form "the lover in the wardrobe", where the PP is a modifier which attributes the property of being located "in the wardrobe" to the individual "lover". Such expressions are interpreted as follows: the space occupied by the localized object (LO) "lover" is situated with respect to a particular spatial region of its reference object (RO) "wardrobe", namely in its interior, referred to by the preposition "in".

3.1. A formal representation language for spatial relations

In order to understand the prepositions' meaning representations presented in section 3.2., I will briefly introduce some formal specification of the representation language. Spatial properties of objects, and the relations between them, are expressed in terms of set theoretic topology. I define spatial objects as sets of spatial points. Spatial prepositions denote the relations between particular subsets of sets of points which are expressed in terms of inclusion and connectedness.

I make the following assumptions: I is a set of concrete objects a, b, c, ... $(P,\mathcal{N})$ is the three-dimensional Euclidian space. $(P,\mathcal{N})$ is a metrical and connected topological space. Let P(a) be the space occupied by an object a. It is defined as follows:

(1) $P(a) = \{p \in P \mid p \text{ is occupied by a}\}$ and $P(a) \subseteq P$

$\forall a \forall b \; a, b \in I \; (P(a) \neq \emptyset \wedge P(b) \neq \emptyset \wedge P(a) = P(b)) \rightarrow a = b$

Definition (1) states that the space P(a) occupied by the object a consists of all points which are occupied by that object. The last clause excludes that two objects may occupy the same space.

There are objects, e.g. cupboards and vases, with respect to which one has to distinguish between the space which is occupied by their material parts and their empty interior, which also belongs to the space these objects occupy. I refer to those spaces with $P_{MAT}(a)$, the set of materially occuppied spatial points, and $P_{EMPTY}(a)$, the set of empty spatial points. They are introduced in (2) and (3) respectively.

(2) $P_{MAT}(a) = \{p \in P \mid p \in P(a) \wedge p \text{ is materially occupied by a}\}$

(3) $P_{EMPTY}(a) = \{p \in P \mid p \in P(a) \wedge p \text{ is not materially occupied by a}\}$

The neighborhood of an object, its vicinity $P_{EXT}(a)$, has to be defined to represent such prepositions, as "near", "by" or "next to".

(4) $P_{EXT}(a) \subseteq C(P(a))$ and $P_{EXT}(a)$ is connected in P

$P_{EXT}(a) = \{p' \in P \mid p' \notin P(a) \wedge \exists p \in P(a) \text{ such that } d(p, p') \leq n_a\}$

As shown in (4), $P_{EXT}(a)$ is a subset of the topological space C(P(a)) which is complementary to P(a) and has a limited extension since another object cannot be localized with respect to object a at arbitrary distances. There is an object and situation specific "region of interaction" ([MJ76]) that confines the object's vicinity. The distance from any exterior point of the object a to an interior point of a should not be smaller than a real number n_a which has to be determined contextually.

Some subsets of an object's exterior have to be defined that are needed for the representation of prepositions as "above", "below" or "by". For the definition of the exterior

above object a, $P_{TOP\text{-}EXT}(a)$, in (6), we need the auxiliary definition in (5). It defines the distance between the set of points occupied by object a and some exterior point p' as the minimum of the distances from any point of P(a) intersected by a vertical projection of p'.

(5) $d_{vert}(P(a), p') = \min_y \{d(y, p') \mid y \in P(a) \cap \pi_{vert}(p') \wedge p' \in P_{EXT}(a)\}$

(6) $P_{TOP\text{-}EXT}(a) = \{p' \in P_{EXT}(a) \mid d_{vert}(P(a), p') > 0\}$

The exterior below object a $P_{BOTTOM\text{-}EXT}(a)$ is defined analogously. The exterior extending horizontally to it, $P_{HOR\text{-}EXT}(a)$, is represented in (7).

(7) $P_{HOR\text{-}EXT}(a) = P_{EXT}(a) \setminus (P_{TOP\text{-}EXT}(a) \cup P_{BOTTOM\text{-}EXT}(a))$

Finally, we need a definition of the object's surface referred to by prepositions such as "on". According to definition (8), the surface consists of all materially occupied points s of an object a, the neighborhood of which contains at least one point p' belonging to the exterior $P_{EXT}(a)$.

(8) $S(a) = \{s \in P_{MAT}(a) \mid \forall N \in \mathcal{N}(s)\ [\exists p' \in N \wedge p' \in P_{EXT}(a)]\}$

With respect to the position in space some subsets of the surface are distinguished: the top, $S_{TOP}(a)$, in (9), the bottom, $S_{BOTTOM}(a)$, which is defined by analogy to (9), and the lateral surface $S_{VERT}(a)$ in (10).

(9) $S_{TOP}(a) = \{s \in P_{MAT}(a) \mid \forall N \in \mathcal{N}(s)\ [\exists p' \in N \wedge p' \in P_{TOP\text{-}EXT}(a)]\}$

(10) $S_{VERT}(a) = S(a) \setminus (S_{TOP}(a) \cup S_{BOTTOM}(a))$

To represent the relations expressed by spatial prepositions I make use of connectedness, represented by the connective "©", and spatial inclusion, represented by the connective "⊏" for complete and "$_p$⊏" for partial inclusion. For definitions, see [BW94a].

3.2. Lexical representations for spatial prepositions

I will use the above defined formal language to describe the meanings of the mentioned topological prepositions in German and English in a strictly monolingual way. I restrict the presentation to some of their meanings that serve as the basis for the demonstration of the translation mapping in the outlined IL model.

Expressions of the type "the lover in the wardrobe" have the semantic sturcture a(prep(b)) which is conceptually interpreted as a spatial relation between two spatial objects and may have an additional functional dimension. Since the conceptual representation should come up with analysis and generation requirements at once, it has to include besides the interpretational part a restrictional part that constrains the use of a certain preposition in a given situation. It consists of spatial and sortal constraints on the localized and reference object. Thus, the lexical entry of a spatial preposition contains the information shown in (12).

(12) a(prep(b)) → spatial and sortal constraints on the localized and reference object ∧
spatial relation ∧ functional relation

The meaning of a preposition is represented as the disjunction of all its possible conceptual interpretations. That means for the analysis: if a certain spatial preposition is used one of the given interpretations including its restrictions is true; and for the generation: if one of the described spatial and functional relations holds and the involved objects fulfill the restrictions the spatial preposition can be applied. For the sake of simplicity I will leave the list of constraints mostly unspecified.

The German preposition "in" ('in') provides two distinct interpretations.

(13) a(in(b)) ↔ (i) [constraints ∧ $P(a) \sqsubset P_{EMPTY}(b)$] ∨
(ii) [constraints ∧ $P(a) \sqsubset P_{MAT}(b)$]

The clause (13i) says that the space occupied by the LO is contained in the empty interior of the RO, cf. (14a). (13ii) accounts for the inclusion of the space which is taken up by the LO in the materially occupied interior of the RO, cf. (14b).

(14a) *die Milch in der Tasse* ('the milk in the cup')
die Besatzung im Flugzeug ('the crew on the plane')
(14b) *der Fisch im Wasser* ('the fish in the water')

The preposition "auf" ('on') has a more complex interpretation, as can be seen in (15):

(15) a(auf(b)) ↔ (i) [constraints ∧ S(a) © S_{TOP}(b) ∧ Supp(b,a)] ∨
(ii) [constraints ∧ S(a) © S_{VERT}(b) ∧ $Supp_{ad}$(b,a)] ∨
(iii) [constraints ∧ b ∈ [auf-idio.use.obj] ∧ P(a) = P_{EMPTY}(b)]

In (15i) the preposition "auf" describes the surface of the LO being in contact with the top surface of the RO, cf. (16a). In general, this spatial configuration implies a support relation between the involved objects. While (15i) is the default reading of "auf" (15ii) covers a more restricted use. Here, the LO has homogeneous contact with the vertical surface of the RO. The use of "auf" for the reference to a lateral surface is restricted to thin surface-like objects and various kinds of substances in the role of the LO, as exemplified in (16b). Here the special support relation of adhesion keeps the LO in such a position. Finally, (15iii) makes explict the idiosyncratic use of "auf" which denotes the inclusion of the LO in the interior of its reference object, cf. (16c). It is restricted to a few objects only which have to be marked in the lexicon, since this use cannot be predicted by the spatial properties of these RO.

(16a) *der Deckel auf dem Topf* ('the lid on the pot')
(16b) *das Etikett auf der Flasche* ('the label on the bottle')
(16c) *die Lampe auf dem Flur* ('the lamp in the hall')

Now let us look at the spatial meaning of the preposition "an" ('on', 'by') in (17):

(17) a(an(b)) ↔ (i) [constraints ∧ $P(a) \sqsubseteq_p S(b)$ ∧ Supp(b,a)] ∨
(ii) [constraints ∧ S(a) © $S_{\neg TOP}$(b) ∧ Supp(b,a)] ∨
(iii) [constraints ∧ P(a) ⊂ $P_{HOR\text{-}EXT}$(b)]

Representation (17i) denotes that the LO is a kind of appendage of the RO which is embedded in the RO's surface, cf. (18a). This use of "an" often goes with a meronymic relation between the involved objects. In (17ii) "an" captures the contact of the LO with a part of the RO's surface that is not the top surface, as shown in (18b). This interpretation partially overlaps with the "auf"-reading in (15ii). If the conditions for the use of "auf" are fulfilled both prepositions can be used. Otherwise only "an" is applicable. Under (17iii) "an" says that the LO is situated in the horizontally extended vicinity of the RO, cf. (18c). To separate the use of "an" from "bei" ('by', 'near'), another preposition for the reference to the exterior of an object, the restriction part of (17iii) contains constraints on the size and boundedness of the reference object.

(18a) *der Griff an der Tür* ('the knob on the door')
(18b) *das Plakat an der Wand* ('the poster on the wall')
(18c) *der Baum am Fluß* ('the tree on the river')

Now I turn to the English topological prepositions. The meaning of "in" is represented in (19).

(19) a(in(b)) ↔ (i) [constraints ∧ P(a) ⊂ P_{EMPTY}(b)] ∨
(ii) [constraints ∧ P(a) ⊂ P_{MAT}(b)]

English "in" captures the same set of interpretations in the spatial domain as the German preposition "in" does, cf. (20a) and (20b).

(20a) *the milk in the cup* *the lamp in the hall*
(20b) *the fish in the water*

As we can see in (21), the English spatial preposition "on" has a wider interpretation.

(21) a(on(b)) ↔ (i) [constraints ∧ S(a) © S(b) ∧ Supp(b,a)] ∨
(ii) [constraints ∧ $P(a) \sqsubseteq_p S(b)$ ∧ Supp(b,a)] ∨
(iii) [constraints ∧ b ∈ [on-idio.use.vehicle] ∧ P(a) = P_{EMPTY}(b)] ∨
(iv) [constraints ∧ P(a) ⊂ $P_{HOR\text{-}EXT}$(b)]

In (21i) "on" expresses the contact of the LO's surface with an arbitrary part of the RO's surface, where the RO supports the LO, cf. (22a). The second representation in (21ii) states that a part of the LO is embedded in the surface of its RO. This is exemplified in (22b). The interpretation in (21iii) captures the idiosyncratic use of "on" for the reference to the interior of a vehicle, cf. (22c). Finally, representation (21iv) indicates that the English preposition "on" is used to situate an object in the neighborhood of ROs which are confined to line-bounded locations, cf. (22d).

(22a) *the lid on the pot* *the lable on the bottle*
(22b) *the knob on the door* *the leaves on the branch*
(22c) *the crew on the plane*
(22d) *the tree on the river*

In the next section the introduced meanings are applied to the proposed IL model.

4. Flexible Interlingua in action: translation of spatial prepositions

Let us return to the modified IL model in section two and apply the ideas outlined there to the translation of topological prepositions. For each of the interpretations of a SL preposition the negotiator has to pick out identical and overlapping representations in the TL lexicon in order to identify the possible TL correspondences and the translation relation between them. From this an analysis strategy is derived.

Comparing the meaning of the considered topological prepositions in German and English, we are faced with equivalence relations as well as cases of generalization and specification. I will exemplify the treatment of these phenomena with the translation of the German prepositions "in" and "auf" and of the English preposition "on".

4.1. The case of equivalence - the translation of the German "in"

The German preposition "in" shares its interpretations with the English preposition "in", cf. (23a,b). Only in the case, where the RO in the TL is used idiosyncratically, the preposition "on" occurs as translation correspondence, cf. (23c).

(23a) *die Milch in der Tasse* *the milk in the cup*
(23b) *der Fisch im Wasser* *the fish in the water*
(23c) *die Besatzung im Flugzeug* *the crew on the plane*

These correspondences are derived in the proposed IL model as shown in (24):

(24i)

SL preposition		conceptual representation	derived TL preposition
a(in(b))	→	constraints ∧ $P(a) \sqsubset P_{EMPTY}(b)$	
		constraints ∧ $P(a) \sqsubset P_{EMPTY}(b)$	→ a(in(b))
		constraints ∧ b ∈ [on-idio.use.vehicle] ∧ $P(a) \sqsubset P_{EMPTY}(b)$	→ a(on(b))

For the interpretation of the LO's inclusion in the empty interior of the RO in (24i) the negotiator finds two TL representations which share this meaning. They are lexicalized by the English preposition "in" or "on" respectively, cf. (23a,c), where the use of "on" is restricted to ROs which are marked for that interpretation in the TL lexicon.

(24ii)

SL preposition		conceptual representation	derived TL preposition
a(in(b))	→	constraints ∧ $P(a) \sqsubset P_{MAT}(b)$	
		constraints ∧ $P(a) \sqsubset P_{MAT}(b)$	→ a(in(b))

The second representation of the German preposition "in" for the LO's inclusion in a materially occupied interior is completely covered by the English preposition "in", cf. (24ii). This is exemplified in (23b).

From this situation the negotiator infers that the set of SL meanings can be handed

over to the generator. No analysis is required since the TL preposition "in" preserves the ambiguity of the SL preposition "in" and the idiosyncratic use of "on" as translation correspondence can be checked in the TL lexicon.

4.2. The case of generalization - the translation of the preposition"auf"

As indicated in (25), the English prepositions "on" and "in" correspond to "auf".

(25a) *der Deckel auf dem Topf* — *the lid on the pot*
(25b) *das Etikett auf der Flasche* — *the label on the bottle*
(25c) *die Lampe auf dem Flur* — *the lamp in the hall*

(25a,b) exemplify the case of TL generalization. While "auf" refers to particular parts of the object's surface, their English counterpart "on" has a wider range of usage. It highlights the contact with any surface of its RO. Using that TL preposition the translation contains less information than in the SL. However, in most cases the loss of information is accomodated by world knowledge, cf. (25a).
In (25c) the idiosyncratic use of "auf" for the inclusion in the RO's interior is lexicalized by the preposition "in" which covers this interpretation by default.

Now let us look at the formal derivation of these translation correspondences. At the first step the negotiator finds the overlapping representations for each of the "auf"-interpretations. Making use of the knowledge about the sets of spatial points, the translation situation is then identified. This is shown in (26).

(26i)

SL preposition		conceptual representation	derived TL preposition
a(auf(b))	→	constraints ∧ $S(a) \copyright S_{TOP}(b) \wedge Supp(b,a)$	
		constraints ∧ $S(a) \copyright S(b) \wedge Supp(b,a)$	→ a(on(b))

With the subset relation $S_{TOP}(b) \subseteq S(b)$ and the given compatibility of the constraint parts, it can be inferred that "on" is a corresponding preposition with a more general meaning, cf. (25a).

(26ii)

SL preposition		conceptual representation	derived TL preposition
a(auf(b))	→	constraints ∧ $S(a) \copyright S_{VERT}(b) \wedge Supp_{ad}(b,a)$	
		constraints ∧ $S(a) \copyright S(b) \wedge Supp(b,a)$	→ a(on(b))

The "auf"-representation in (26ii), exemplified in (25b), subsumes the representation of "on" because of the subset relation $S_{VERT}(b) \subseteq S(b)$ and the fact that the adhesion implies support. The restrictions on the "auf"-interpretation are contained in those of the "on"-reading.

(26iii)

SL preposition		conceptual representation	derived TL preposition
a(auf(b))	→	constraints ∧ b ∈ [auf-idio.use.obj] ∧ $P(a) \sqsubset P_{EMPTY}(b)$	
		constraints ∧ $P(a) \sqsubset P_{EMPTY}(b)$	→ a(in(b))
		constraints ∧ $P(a) \sqsubset P_{EMPTY}(b)$ ∧ b ∈ [on-idio.use.vehicle]	→ a(on(b))

If the preposition "auf" is used idiosyncratically, as in (25c), which can be easily identified by the markedness of the German RO since the idiosyncratic use is given preference, two identical representations are found, cf.(26iii). They are lexicalized by "in" or "on" respectively. Which of the two prepositions is the appropriate one is determined by the lexical restriction on the English RO.

Summarizing, the preposition "auf" can be translated without analysis since the idiosyncratic use in (26iii) can be handled by lexical markedness and the translation mapping of (26i) and (26ii) is justified because of the recognized generalization which allows the mapping by implication.

4.3. The case of specification - the translation of the preposition "on"

The translation of the preposition "on" is rather complicated. On the one hand this preposition is ambiguous with respect to the SL, cf. (21), and on the other the TL

makes a lexical distinction which is not predictable from the SL perspective. While the English preposition "on" captures primarily the contact with any part of the RO's surface, we have to choose between the more specific prepositions "an" and "auf" in German, as shown in (27a,b). In this case the translation renders the information more explicit than in the SL. The "on"-reading of the LO's embedding in the surface of the RO in (27d) is shared by the preposition "an". Moreover, "on" provides the interpretation of the LO's inclusion in the RO's interior which is captured by the German "in", as in (27e), and its localization in the RO's vicinity covered in the TL by the preposition "an", cf. (27f).

(27a)	*the lid on the pot*	*der Deckel auf dem Topf*
(27b)	*the poster on the wall*	*das Plakat an/auf der Wand*
(27d)	*the knob on the door*	*der Griff an der Tür*
(27e)	*the crew on the plane*	*die Besatzung im Flugzeug*
(27f)	*the tree on the river*	*der Baum am Fluß*

These translation correspondences are established as follows. For the particular meanings of the preposition "on", introduced in (21), the negotiator picks out the corresponding TL representations listed in (28).

(28i)	SL preposition		conceptual representation	derived TL preposition
	a(on(b))	→	constraints ∧ S(a) © S(b) ∧ Supp(b,a)	
			constraints ∧ S(a) © S_{TOP}(b) ∧ Supp(b,a)	→ a(auf(b))
			constraints ∧ S(a) © $S_{\neg TOP}$(b) ∧ Supp(b,a)	→ a(an(b))
			constraints ∧ S(a) © S_{VERT}(b) ∧ $Supp_{ad}$(b,a)	→ a(auf(b))

If "on" denotes the contact with any part of the RO's surface, as in (28i), three overlapping TL representations are found. They are identified by the superset relations S(b) ⊇ S_{TOP}(b), S(b) ⊇ $S_{\neg TOP}$(b) and S(b) ⊇ S_{VERT}(b) respectively, cf. (27a,b). All these TL representations are more specific than in the SL, i.e. the TL makes a conceptual distinction that is covered by different prepositions. To choose the appropriate one the analysis has to identify which of the TL relations is relevant. Thus, the depth of the analysis is predicted by the TL specification.

(28ii)	SL preposition		conceptual representation	derived TL preposition
	a(on(b))	→	constraints ∧ P(a) $_p$⊏ S(b) ∧ Supp(b,a)	
			constraints ∧ P(a) $_p$⊏ S(b) ∧ Supp(b,a)	→ a(an(b))

If "on" is used to express the embedding of the LO in the RO's surface, as in (28ii), an identical representation is picked out. It is lexicalized in the TL by the preposition "an", cf. (27d).

(28iii)	SL preposition		conceptual representation	derived TL preposition
	a(on(b))	→	constraints ∧ b ∈ [on-idio.use.vehicle] ∧ P(a) ⊏ P_{EMPTY}(b)	
			constraints ∧ P(a) ⊏ P_{EMPTY}(b)	→ a(in(b))
			constraints ∧ b ∈ [auf-idio.use.obj.] ∧ P(a) ⊏ P_{EMPTY}(b)	→ a(auf(b))

If "on" is used to denote the inclusion of the LO in a vehicle two corresponding TL representations are recognized. Here, the lexicalization in the TL depends on the markedness of the German RO, cf. (27e).

(28iv)	SL preposition		conceptual representation	derived TL preposition
	a(on(b))	→	constraints ∧ P(a) ⊏ $P_{HOR\text{-}EXT}$(b)	
			constraints ∧ P(a) ⊏ $P_{HOR\text{-}EXT}$(b)	→ a(an(b))

For the "on"-reading of localizing an object in the vicinity of the RO in (28iv) an identical TL representation is found where the constraints of application are less specific than in the SL so that "on" can be substituted by "an".

Summarizing, from the contrastive situation for the preposition "on" can be concluded that in (28i) an analysis according to the TL specification is required. Since

other readings of "on" are also lexicalized by different prepositions the analysis has to be planned in a global way, i.e. the disambiguation strategy is determined with respect to the effort for identifying of the appropriate reading. Since the focus of this paper lies on the demonstration of the negotiator's work I cannot go into details of the disambiguation which calls for a great amount of world and contextual knowledge, see [BW94a].

5. Summary

In this paper I looked at problems of the interlingual representation of lexical items without coinciding meanings in different languages. I introduced the idea of a flexible Interlingua representation which fits the actual contrastive situation in two ways. On the one hand, it contains all meanings of a SL lexical item which are lexicalized in the TL by the same word, i.e. ambiguities which are preserved in the languages involved are recognized in advance and the analysis is restricted to the necessary depth. On the other hand, in the case of conceptual divergences it corresponds either to the more general or to the more specific TL representation. Modelling a negotiation device that establishes this kind of flexible IL representations, we get by with strictly monolingual lexical representations which are applicable to both analysis and generation.

Since the outlined approach provides a tool both for recognizing ambiguities which are preserved by the TL and for handling generalization and specification gaps, it can be applied to the interlingual translation of a wide range of nouns and verbs which are ambiguous either in the SL or with respect to the TL, see [BW94b]. The lexical representations of such lexical items can also be viewed as conceptual representations that are organized in a multi-lingual concept hierarchy from which the relations of super- and subordination between SL and TL concepts can be derived.

6. References

[BMR91] J. Barnett, I. Mani, E. Rich, C. Aone, K. Knight, J.C. Martinez: Capturing Language-Specific Semantic Distinctions in Interlingua-Based MT. In: Proceedings of the Machine Translation Summit III, 25-32, Washington D.C., 1991.

[BW94a] B. Buschbeck-Wolf: Topologische Präpositionen und ihre Verarbeitung in der Maschinellen Übersetzung. Dissertation, Manuscript, Stuttgart, 1994.

[BW94b] B. Buschbeck-Wolf: Lexicalization and conceptualization from a contrastive point of view. To appear in: Proceedings of the Workshop: "Semantisches und konzeptuelles Wissen", Berlin, 21.-23.4.1994.

[BEH94] B. Buschbeck-Wolf, M. Egg, M. Herweg: Desambiguierungsverfahren für die konzeptuelle Interpretation polysemer Präpositionen. In preparation, IBM Heidelberg, 1994.

[HS92] W. J. Hutchins, H. L. Somers: An Introduction to Machine Translation. Cambridge: Academic Press, Cambridge, 1992.

[KOP91] M. Kameyama, R. Ochitani, S. Peters: Resolving Translation Mismatches With Information Flow. In: Proceedings of the 29th Annual Conference of the ACL, 193-201, Berkeley, 1991.

[KGN91] M. Kay, J. M. Gawron, P. Norvig: Verbmobil: A Translation System for Face to Face Dialog. Manuscript, Stanford University, 1991.

[MJ76] G.A. Miller, P.N. Johnson-Laird: Language and Perception. Cambridge University Press, 1976.