Lexical Allocation in Interlingua-based Machine Translation of Spatial Expressions*

Clare R. Voss and **Bonnie J. Dorr** and **M. Ülkü Şencan**
University of Maryland
College Park, MD 20742
{voss, bonnie, ulku}@cs.umd.edu

Abstract

Given a spatial expression, or its computational semantic form, how is the expression's spatial semantics to be allocated *lexically,* i.e., among entries in the lexicon? In interlingua-based machine translation (MT) research, *lexical allocation* is the problem of *allocating* or subdividing a linguistic expression's full interlingual (IL) structure into the substructures that are *lexical* IL forms, i.e., in the lexicon. Here we present our work developing IL forms and an *IL lexicon* for translating English spatial expressions into Turkish. We examine several co-occurrence patterns between motion verbs (spatial placement and displacement) and directional adpositions (particles in English, postpositions in Turkish) and the lexical allocation of spatial vectors in these patterns.

1 Introduction

In this paper we report on our current research developing computational forms for the interlingua-based (IL) machine translation (MT) of spatial expressions. We frame this research in terms of the following problem in *lexical allocation*:

> Given a spatial expression, or its computational semantic form, how is the expression's spatial semantics to be allocated *lexically,*[1] i.e., among entries in the lexicon?

In IL-based MT the allocation problem refers to subdividing a linguistic expression's full interlingual (IL) structure into the substructures, each of which corresponds to a lexicon entry's IL form, i.e., a *lexical* IL form. The problem appears at two phases:

1. During construction of MT lexicon entries (at MT definition-time): given a spatial expression, how should the expression's spatial semantics be represented in an IL form and then be subdivided among the source language (SL) lexical elements present in the expression?
2. During lexical selection (at MT runtime): given an IL form derived from a spatial expression, how should that IL form be subdivided among the target language (TL) lexical entries available in the MT system?

In this paper we focus on the lexical allocation of spatial directional information in the translation of expressions from English to Turkish. Turkish postpositions and case markings help differentiate directions from goal-directed paths, giving us insights into the components of spatial semantics needed for translations to target languages where different lexical items are selected based on this distinction. We present expressions whose verb phrases contain both a motion verb (spatial placement or displacement) and a directional adposition (particles in English and postpositions in Turkish)[2] and yet display distinct surface co-occurrence patterns. We describe our work developing IL forms and an *IL lexicon* for these two types of lexical items and the spatial vectors appearing in their co-occurrence patterns.

2 Spatial Semantics in IL Forms

We use the following terms to distinguish among different types of spatial information in the IL form of a linguistic expression:[3]

- *Spatial object-functions* include the geometric description functions of [Herskovits, 1986][4] and the

*This research was supported, in part, by the Army Research Office under contract DAAL03-91-C-0034 through Battelle Corporation, by the National Science Foundation under grants NYI IRI-9357731, NSF/CNRS INT-9314583, by the Alfred P. Sloan Research Fellow Award BR3336, and by the Army Research Institute under contract MDA-903-92-R-0035 through Microelectronics and Design, Inc.

[1] *Lexical* is not used to distinguish *lexical categories* from *functional categories;* rather it refers to meaning components in the lexicon.

[2] Here *adpositions* includes prepositions, postpositions, and particles.

[3] This set is not comprehensive with respect to natural language spatial semantics [Dorr and Voss, 1993]. It is beyond the scope of our paper to address the complexities of functional and pragmatic levels of representation. For an approach that addresses these levels, including the geometric one we examine here, see [Aurnague and Vieu, 1993].

[4] Herskovits identifies several object idealizations, parts, forms, volumes, axes and projections as *geometric description* (GD) functions on an object O at a time i.

schematization of objects of [Jackendoff, 1991].[5] These may take an object and yield a *spatial entity* (eg., place), or they may apply recursively to a spatial entity. For example, *the top of the hill* is a spatial entity where the function *top of* has an object, *hill,* as its argument.[6]

- *Spatial predicates* capture the positions of two entities relative to each other.[7] That relative positioning is assessed at a moment in time. These spatial configurations are locational and directional relations among entities situated in the real, 3-dimensional, physical world, such as when one entity is *above* another entity.[8]
- *Spatial situations* are events and states in the real, 3-dimensional, physical world, that are assessed spatio-temporally, such as those involving motion, position, or containing a spatial configuration.[9]

This framework leaves as an open question what the representation is for a spatial *vector,* i.e., a *directed* one-dimensional axis with an origin but no other reference points. Vectors are clearly geometric abstractions. However they also appear implicitly in ordinary, natural languages. Consider, for example, the direction *up,* a lexicalization of a vector, and words with an *up* prefix sharing that semantics. When we say that something is located *upstate* or *uphill,* then we know it is in *the upper part of* the state or toward or at *the top of* the hill. These *up*s behave like *spatial object-functions* yielding the uppermost portion of a spatial entity, a place. The origin of the *up* vector is established by an entity (state or hill) and its direction is a function on that entity.

On the other hand, when we say that a canoe headed *upriver* or a bird headed *upwind,* then we only know that they headed in a direction established with reference to the river or the wind. These *up*s behave like *spatial predicates* yielding a path. The predicator is the direction of this *up* vector. One argument is the entity following the direction[10] and the second argument is a reference path. Finally we only note here that it is beyond the scope of this paper to extend the discussion to vector directions and spatial situations.[11]

In this paper we examine these vector representation issues for IL forms for the English word *up* and verbs of spatial motion. We ask what lexical IL forms are needed to capture their spatial semantics in translating from English to Turkish.

3 Data Set

In this section we examine the co-occurrence patterns of spatial displacement verbs and directional adpositions together in verb phrases. We first look at the patterns in English expressions and then at translations of these expressions into Turkish. Our goal is to identify expressions close in overall sentential meaning but whose individual verbs vary in their co-occurrence patterns with adpositions. Accounting for this contrast — close sentential semantics but distinct lexical co-occurrence patterns — requires allocating distinct lexical semantics to the verbs.

3.1 Patterns in English

In the expressions below, '*' and'*/?' mark sentences judged ungrammatical and questionable (i.e., an unsure judgment) by native English speakers, repsectively. '(uM)' marks grammatical sentences whose verbs have an implicit upward motion (placement or displacement).

Yesterday we watched as the crane operator .....

(1) (i) (uM) elevated the new railroad tracks.
(ii) *elevated **up** the new railroad tracks.
(iii) *elevated **down** the new railroad tracks.

(2) (i) (uM) lifted the new railroad tracks.
(ii) (uM) lifted **up** the new railroad tracks.
(iiI) */?lifted **down** the new railroad tracks.

(3) (i) *put the new railroad tracks.
(ii) put **up** the new railroad tracks.
(ii) put **down** the new railroad tracks.

(4) (i) moved the new railroad tracks.
(ii) moved **up** the new railroad tracks.
(ii) moved **down** the new railroad tracks.

Summary of Patterns for English

The table below summarizes the co-occurrence patterns from the sentences above. The '*' and '?' are used as above. For example, the combination of *elevate* and *up* is ungrammatical. We introduce here a '+' and '(+)' for grammatical combinations where the argument (not

[5] Jackendoff's including and extracting functions are not all strictly spatial. His schematization of dimensionality is spatial and extends to states and events.

[6] [Reyero-Sans and Tsujii, 1994] have a similar type they identify as *structural terms,* in contrast to *relational terms.* [Aurnague and Vieu, 1993], per A. Borillo (1988), use the term *internal localization nouns* for lexical elements that specify different portions of an entity.

[7] We use *predicate* here in the logico-mathematical sense of the term, not its linguistic usage, as a verb phrase consisting of a verb and its internal argument(s). Unless otherwise noted, we use *predicate* for the full structure of a predicator and its arguments.

[8] Most typical examples of spatial predicates in the literature are prepositions (eg., see [Retz-Schmidt, 1988] for an excellent survey). For a presentation that goes beyond the syntactic category of prepositions, see [Heine *et al.*, 1991].

[9] There may also exist another category distinct from, or a subset of, spatial situations. It carries spatial configuration and semantic aspect, but not temporal location, which it derives from another phrase. For example, in *he came into the room, running,* the action of *running* is contemporaneous with the *come* motion, deriving its temporal location (past) and directionality (*into the room*) from the main phrase in the expression.

[10] This is often referred to as the figure [Talmy, 1985].

[11] Verbs such as *upend* and *uproot* provide evidence that *up* may appear as a spatial-object function and a spatial predicate, respectively, within a spatial verb. However the status of a third type of *up* vector, distinct from these two and within a spatial situation, is not clear.

the specific direction) is obligatory and optional, respectively.

1. Obligatorily Lexically Implicit
   eg., *elevate* [* UP] [* DOWN]
   In this class verbs have an inherent spatial direction in its meaning and this direction cannot be explicitly lexicalized in combination with the verb.
2. Lexically Implicit and Optionally Explicit
   eg., *lift* [(+) UP] [*/? DOWN]
   Verbs in this class have an inherent default direction and this direction, unlike in the previous class, can be made explicit.
3. Obligatorily Lexically Explicit
   eg., *put* [+ UP] [+ DOWN]
   A verb in this class has no inherent direction, yet it requires an explicit locational argument.
4. Optionally Lexically Explicit
   eg., *move* [(+) UP] [(+) DOWN]
   Verbs in this class have no inherent specific direction, only a reading of spatial displacement in the event.

3.2 Patterns in Turkish

Here we present Turkish translations of the English sentences from the previous subsection and then summarize the direction classes for the verbs involved. Turkish provides an interesting MT test case for the spatial semantics of English *up,* which may be (i) a vector equivalent in meaning to *upward* or (ii) a goal-marked path equivalent in meaning to *to* or *toward some place that is up (higher/further along.)* In the sentences below the English word *up* may translate into the Turkish *yukarı* (root form) as *upward* or *to a place that is up/higher*. In the former case of the vector direction *upward, yukarı* is used. We will refer back to this adverbial sense as *y-1.* In the latter case of a path to or toward some goal,[12] we find two forms: (i) *yukarı* or (ii) *yukarıya.* We will refer back to these related nominal senses as *y-2* and *y-3.*[13]

E: Yesterday we watched as the crane operator .....[14]
T: Dün, vinç operatörünün ..... seyrettik.
'Yesterday, crane operator ..... watched.'

(5) (i) E: **elevated** the new railroad tracks.
T: yeni demiryolu raylarını **yükseltişini**
'new railroad tracks elevated'

(ii) E: elevated ***up**/***down** the new railroad tracks.
T: yeni demiryolu raylarını ***yukarı** /***aşağı** yükseltişini [15]

(6) (i) E: **lifted** the new railroad tracks.
T: yeni demiryolu raylarını **kaldırışını**

(ii) E: lifted **up**/?***down** the new railroad tracks.
T: yeni demiryolu raylarını (**yukarı/yukarıya**) / *(**aşağı/aşağıya**) kaldırışını

(7) (i) E: ***put** the new railroad tracks.
T: yeni demiryolu raylarını **koyuşunu**
'new railroad tracks **placed**'

(ii) E: put **up/down** the new railroad tracks.
T: yeni demiryolu raylarını
(**yukarı/yukarıya**) / (**aşağı/aşağıya**) koyuşunu
'new railroad tracks **up/down** placed'

(8) (i) E: **-moved** the new railroad tracks.
T: yeni demiryolu raylarını **hareket ettirişini**
'new railroad tracks **motion do-caused**' [16]

(ii) E: moved **up/down** the new railroad tracks.
T: yeni demiryolu raylarını (**yukarı/yukarıya** (**doğru**)) / (**aşağı/aşağıya** (**doğru**)) hareket ettirişini[17]

It should be noted that above in sentences (6ii) and (8ii), the translation of *up* to *yukarı* is ambiguous, meaning either *y-1* or *y-2,* as described above. However, in sentence (7ii), only the meaning *y-2* is acceptable. That is, the *y-1* interpretation of *yukarı* in (7ii) is strictly ruled out.

Summary of Patterns in Turkish

In the summary table below, the same marking conventions as above in section 3.1 are used. We see here Turkish co-occurrence patterns for classes 1,2, and 4 equivalent to those in section 3.1 for English.

1. Obligatorily Lexically Implicit
   Turkish: *yükselt* [* YUKARI] [* AŞAĞI]
   'elevate' [* UP] [* DOWN]
2. Lexically Implicit and Optionally Explicit
   Turkish: *kaldır* [(+) YUKARI] [* AŞAĞI]
   'lift' [(+) UP] [* DOWN]
3. Obligatorily Lexically Explicit
   Turkish: closest translation of *put* is *koy* in next class
4. Optionally Lexically Explicit
   Turkish: *koy* [(+) YUKARI] [(+) AŞAĞI]
   'place' [(+) UP] [(+) DOWN]
   Turkish: *hareket ettir* [(+) YUKARI] [(+) AŞAĞI]
   'move' [(+) UP] [(+) DOWN]

3.3 Lexical Allocation in the Data Set

We have limited our presentation here to these verb-adposition distribution classes, although there are others.[18] Our aim has not been to provide a complete enumeration of verb classes and verb-adposition alternations.[19] Rather the goal has been to identify expressions close in overall sentential meaning but whose individual verbs vary in their co-occurrence patterns with adpositions. The sentences thus provide a natural test set for different solutions to the lexical allocation

[12] In general the *to* meaning is stronger than the *toward.*

[13] This distinction is slippery indeed, if not elusive, for non-Turkish speakers. The *-a* case suffix on *yukarıya* marks it as a noun. When *yukarı* is read with sense *y-2,* it is a noun, like *y-3,* without an overt case suffix.

[14] We use E and T to designate the original English sentence and the Turkish translation of E. An extra line with the word-for-word English translation of T preserving Turkish word order may also appear in single quote marks as needed.

[15] Translation of this ungrammatical sentence was based on sentence (5i) and inserting the spatial postposition.

[16] This causative sense of *move* is distinct from the reflexive (self-caused) sense which is also *hareket et* in that it generally refers to the start of the motion or a slight movement.

[17] Adding *doğru* forces a goal-directed path reading, suggesting that a *y-2* reading of *yukarı* is possible.

[18] Consider the verb *pick.* We can *pick up* a box but not *pick down* a box. Furthermore if we merely *pick* a box, the verb is possessional rather than spatial.

[19] [Levin, 1993] provides such an extensive list for English, for example. [Talmy, 1985] examines related issues cross-linguistically.

problem, i.e., subdividing sentential meaning in distinct ways for each verb-adposition pair.

4 Allocation and MT Lexicon Construction

This section examines the problem of lexical allocation during MT lexicon construction, before MT runtime.[20] We use *extralexical operations*[21] to capture the compositional properties of lexical elements in the sentences in section 3. After sketching out these operations, we describe our basic lexical IL forms and the annotations on these forms for computing the operations. We end this section with a discussion of additions needed to handle the Turkish data with our IL forms.

4.1 Extralexical Operations

The *extralexical operations* we describe here serve as a bridge from the linguistic data classes in section 3 above to the specific computational approach, annotations on IL forms, to be spelled out further in 4.2 below. With respect to the verbs in the last section, we say that the IL form for the verb and the IL form for *up* are related as follows:

- Obligatorily Lexically Implicit
  elevate: its IL form **blocks** attachment with *up*
  an IL form for *up* is internal to the IL form for *elevate*
- Lexically Implicit and Optionally Explicit
  lift: its IL form may **overlap** the attachment of *up*
  an IL form for *up* is internal to the IL form for *lift*
- Obligatorily Lexically Explicit
  put: its IL form must be **filled** in its locational argument position, no IL form for *up* internal to the IL form for *put*
- Optionally Lexically Explicit
  move: its IL form may be **filled** by attachment of *up*
  no IL form for *up* internal to the IL form for *move*

More generally, given an IL form for a spatial expression, we specify that the IL form allocated for the verb, call it V, and the IL form allocated for *up,* call it P, may be related in one of the following ways:

block: V blocks P from attaching to any node in V
overlap: V permits P to attach and share subtree rooted at a non-leaf node in V
fill: V permits P to attach at a leaf node in V

4.2 Computational Forms

Our approach in developing IL forms derives from the lexical conceptual structures (LCSs) of Jackendoff [1983; 1990; 1991].[22] The LCS framework consists of three independent subsystems: fields, conceptual constituents, and boundedness and aggregation properties. Only the first two are currently a part of our IL forms, as shown in figure 1. The LCS fields (spatial, temporal, possessional, identificational, and others) are motivated by well-known observations of *lexical parallelism,* where the same lexical item has parallel or related meanings in other semantic fields. (We have omitted the *-LOC* field suffix in diagrams of our IL forms, given that all IL forms are for spatial expressions only.)

The conceptual constituents in the second LCS subsystem are variants on predicate-argument structures. They are typed by one of a small set of ontological categories (Thing, State, Event, Place, Path, Property, and Amount) and the internal structure of each constituent decomposes into another conceptual constituent. The predicate primitives are subscripted by field in addition to being typed by category. Although these primitives do not appear in all the fields, Jackendoff does make the claim that the constituent structures do generalize across fields. In particular, he adapts a *localist* view, claiming that the formalism for encoding constituents in the spatial field, at some level of abstraction, generalizes to other fields.[23] At the end of this paper we present a few examples that support this approach — we show that the distinction we build into our IL forms for the spatial field extends to other fields as well.

In the lexical IL forms of figure 1, two types of annotations appear on leaf and non-leaf nodes in order to implement the extralexical operations sketched above. The '*' marker at a node indicates that an obligatory attachment by another IL form occurs at that site. The '(*)' marker indicates an optional IL form attachment may occur at that site. The relation between the distribution classes, the extralexical operations, and the markings on the IL forms appears in the table below.

Class	Extralexical Op.	Marker	Att. Site
Oblig. Lex. Implicit	block	none	not applic.
Lex. Impl. & Opt. Expl.	overlap	(*)	internal
Oblig. Lex. Explicit	fill	*	external
Optionally Lex. Explicit	fill	(*)	external

The annotations on lexical IL forms are used at MT runtime during the analysis and generation phase to guide, respectively, the composition and decomposition of the full IL form corresponding to the input spatial expression. The annotations are language-specific and are removed from the fully composed IL form when the analysis phase is complete.[24]

[20] MT lexicon entries can also be created at runtime, eg., [Onyshkevych and Nirenburg, 1994].

[21] We define an IL's syntax in terms of (i) lexical IL forms within language-specific lexicons and (ii) algorithms for creating and decomposing the instantiated *pivot* representation [Voss and Dorr, 1995]). The *extralexical operations* are an abbreviated version of (ii).

[22] See [Dorr, 1993] for details of a MT system whose IL is LCS-derived. Recently the LCS framework has been used by others for French, eg., [Pugeault *et al.*, 1994; Verrière, 1994].

[23] A localist, or localist-related, approach is by no means unique to Jackendoff. See, for example, among many others, [Anderson, 1971; Heine *et al.*, 1991; Langacker, 1987].

[24] The details of these algorithms are discussed in [Dorr, 1993].

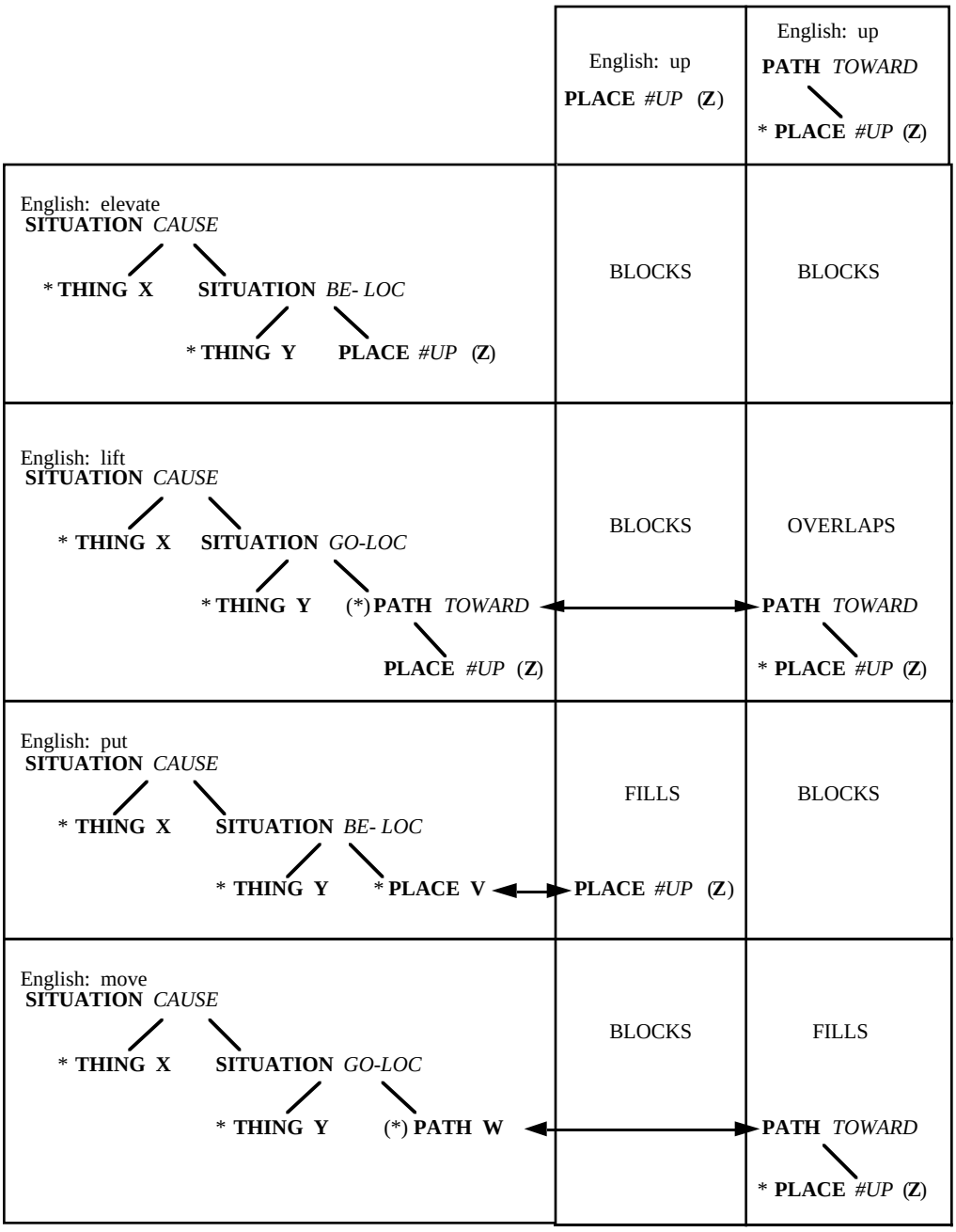

Figure 1: IL forms for English Verbs and *up*

4.3 Lexical Construction for Turkish Data

In the summary tables of section 3, we saw that for three of the four classes, the Turkish verb used to translate the English verb had the same distribution pattern as its English counterpart. The one case where this did not occur required further examination. As noted in that section, the Turkish verb *koy* that comes the closest to English word *put* cannot take as its argument the *y-1,* or simple vector sense of *up.* Although it readily takes Place type arguments and will not take a *y-1* Path as an argument, we discovered that it would accept *y-2* and *y-3,* the goal-marked Path senses, as arguments. This was a surprising fact: why would one Path be acceptable to a verb that generally takes a Place argument while another Path was not? It turned out the goal-marked Path sense was acceptable only under the coerced interpretation that these were Places.[25] That is, for the Turkish equivalent of *put,* simple vectors are apparently not coercible into Places, whereas goal-marked Paths are.

In order to capture this contrast, we have distinct IL forms for simple vectors and goal-marked paths. We hypothesize that *koy,* in looking for a Place as its argument, will accept and coerce a goal-marked path into a Place, but not a simple vector, precisely because the former contains a Place whereas the latter does not. In other words, coercion to a Place is an operation that selects for boundable entities and thus rejects simple vectors. The lexical IL forms to handle these data appear in boxes in figure 2. In particular, a new Path predicate was needed for the simple vector sense of *up.* Previously we had only the Path predicate *TOWARD* in combination with a Place ♯*UP*, consistent with Jackendoff's limited set of five Path primitives. As a side effect of this change, the IL syntax now also allows for new situation predicates formed with this simple vector Path type.

5 Allocation During Lexical Selection

The *allocation* analysis described above occurs at MT lexicon definition-time, when a researcher decides what IL forms to allocate to the lexical entries.[26] Here we describe our work building *IL lexicons* that will be used for lexical selection at MT generation runtime. In an IL-based MT system the process of lexical selection is one part of the generation phase that follows the construction of a *pivot* IL form from the input source language expression. The pivot IL form must be subdivided into IL forms that correspond to lexical IL forms in the target language's lexicon. We split the selection process into several components. Of interest here is the one that, given a subpart of the pivot IL form, will find a range of TL lexical items whose IL forms cover or approximate that given subpart. Following [DiMarco *et al.*, 1993],

[25]For comparison, consider the English sentence *he lives through the tunnel.* This means he lives somewhere that is through, i.e., beyond the tunnel. Here the verb *live,* which requires a Place as its argument type, coerces the Path *through the tunnel* into a Place by treating it as embedded within an indefinite relative clause *somewhere that is through the tunnel.*

[26]The allocation operations execute at runtime, although they originate in the IL forms created at definition-time.

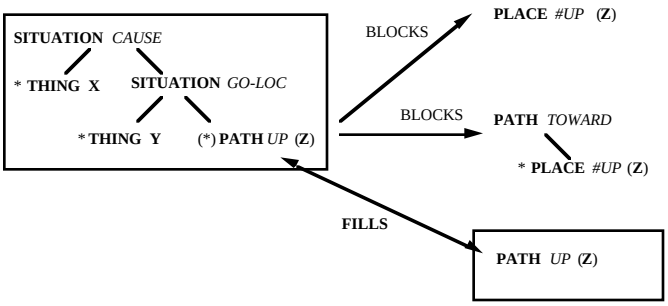

Figure 2: Additional IL forms

we call this component a *lexical option finder.* Our approach has been to bypass direct TL lexicon searches and instead build an *IL lexicon* for the finder to use.

5.1 IL Lexicons

Our IL lexicon is a hierarchical data structure that organizes the space of IL forms extracted from the MT system's lexicons.[27] Its original purposes were (i) to structure the search space of lexical IL forms for the lexical option finder and (ii) to provide a reverse-index into language-specific MT lexicons. We note here in passing that the IL lexicon also *grounds* the IL forms by coindexing them each with lexical items from one or more natural language.

The nodes of an IL lexicon are IL forms corresponding to at least one lexical item's entry in one of the MT lexicons. The nodes are connected by links based on the structural properties of the lexical IL form identifying that node.[28] In figures 3 and 4, the bold arrows stand for structural subsumption relations between node identifiers. The dashed arrows are *reduction links* connecting one lexical IL form to another that is one of its substructures. Attached to each node are reverse-index pointers to its lexicalizations, i.e., the lexical entries in the MT lexicons in the system. That is, the nodes are not lexical items per se, but rather concepts defined only in terms of an IL form identifier without language-specific annotations (such as the *-markers described earlier).[29]

[27]We use the phrase *MT lexicons* to refer to the natural language-specific lexicons in MT systems that are organized by language-specific entries. In an earlier description of this work, we identified the IL lexicon as a *concept-based lexicon.* [Dorr *et al.*, 1994])

[28]The nodes in our figures do not have their LOC field marked.

[29]While this approach resembles that taken by [Wu and Palmer, 1994], who address related issues in lexical selection, our IL lexicon is structured by the syntax of our IL while their hierarchy is defined in terms of concepts that encode a multi-part meaning representation.

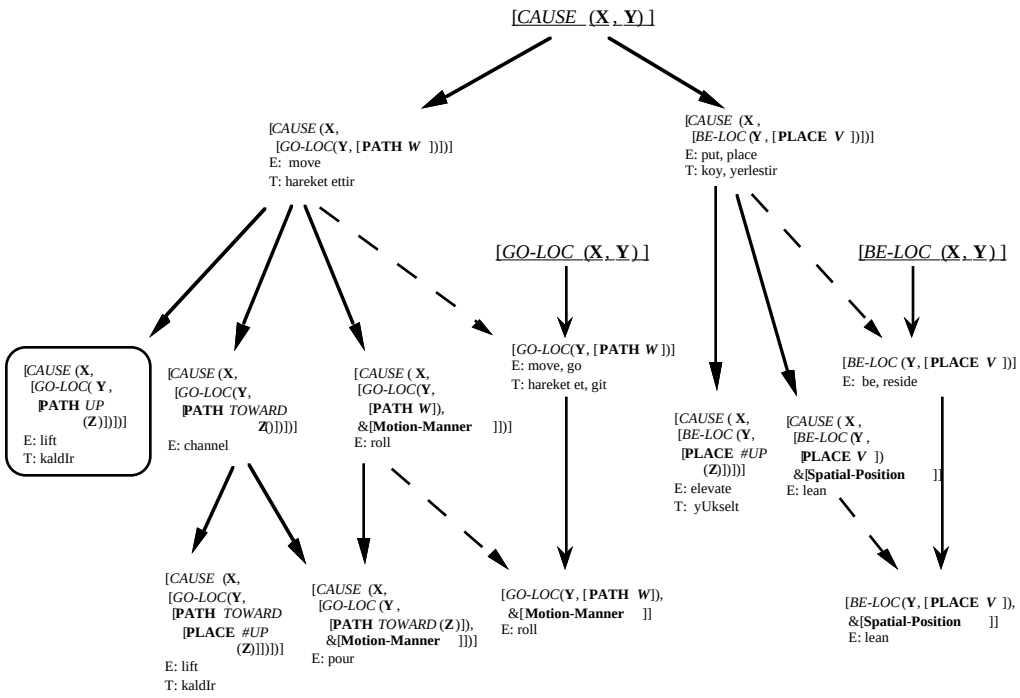

Figure 3: Subgraph of *IL Lexicon* for Spatial Situations

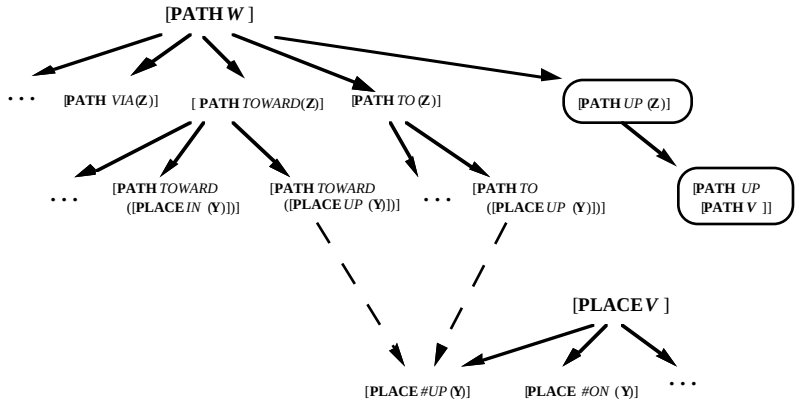

Figure 4: Subgraph of *IL Lexicon* for Spatial Predicates

The lexical option finder traverses the IL lexicon following these links, using nodes annotated with TL lexicalizations to generate the range of lexical items for other components in the lexical selection process.[30] Thus far in our IL lexicon the nodes have been lexical IL forms with a function- or predicate-argument structure. We have relied on the ontology in a knowledge representation system for structuring relations among objects identified by simple IL forms.

5.2 Lexical Selection With Turkish Data

In the IL lexicon figures 3 and 4, the IL forms added in accounting for the distinction between simple vectors and goal-marked paths are boxed. The Turkish simple vector *up* labelled earlier as *y-1* is a lexicalization of [**PATH** *UP*(Z)]. The other Turkish senses of *up, y-2* and *y-3,* are lexicalizations of paths *TOWARD* and *TO* with the place argument ♯*UP*.

6 Conclusion and Future Work

This paper addresses the general problem of lexical allocation in the IL-based MT in conjunction with the specific question of how to represent spatial directions in LCS-derived IL forms. With respect to MT research, we show that, in our IL-based system, the two operationally distinct phases of lexical construction (pre-runtime) and lexical selection (runtime) are implicitly interdependent. Within our framework, these both are problems in lexical allocation and depend on the syntax of the IL. In particular, the IL syntax formalism defines (i) the structures for the spatial information to be allocated to lexical items and (ii) the search space of the IL lexicon to be traversed during lexical selection. Our next step is to build support tools to specify a lexicalized IL grammar and guide the building of lexical IL forms and an IL lexicon.

With respect to spatial expressions, we have pursued the hypothesis that vectors lexicalize in two ways, as spatial object-functions and as spatial predicates. We use these to distinguish (i) goal-marked vectors from (ii) simple vectors in our lexical IL forms, a contrast needed for translation into Turkish.[31] In future work we will expand our data set in order to test this hypothesis further. We note in concluding that our work remains consistent with the localist framework: both of our Path interpretations of *up* can be extended to non-spatial fields, as in the sentences below.

(i) temporal:
They moved up the deadline.
They advanced/delayed the deadline.
identificational:
Her temperature went up (and stayed there).
Her temperature peaked.

(ii) temporal:
He sped up the car.
He accelerated the car.
identificational:
The rocket went up (and up).
The rocket rose/soared.

[30] Further details are presented in [Dorr *et al.*, 1994].

[31] The two vector forms are: (i) [**PATH** *TOWARD* [**PLACE** ♯*UP*(x)]] with the spatial object-function ♯*UP*, and (ii) [**PATH** *UP-LOC*(z)] where the spatial predicate is *UP-LOC*.

References

[Anderson, 1971] J. Anderson. *The Grammar of Case: Towards a Localist Theory.* Cambridge University Press, Cambridge, England, 1971.

[Aurnague and Vieu, 1993] M. Aurnague and L. Vieu. "A Three-Level Approach to the Semantics of Space". In C. Zelinsky-Wibbelt, editor, *The Semantics of Prepositions: From Mental Processing to Natural Language Processing.* Mouton de Gruyter, Berlin, Germany, 1993.

[DiMarco *et al.*, 1993] C. DiMarco, G. Hirst, and M. Stede. "The Semantic and Stylistic Differentiation of Synonyms and Near-Synonyms". In *Working Notes for the AAAI Spring Symposium on Building Lexicons for Machine Translation*, Technical Report SS-93-02, pages 114–121, Stanford University, CA, 1993.

[Dorr and Voss, 1993] B. Dorr and C. Voss. "Machine Translation of Spatial Expressions: Defining the Relation between an Interlingua and a Knowledge Representation System". In *Proceedings of the AAAI*, pages 374–379, Washington, DC, 1993.

[Dorr *et al.*, 1994] B. Dorr, C. Voss, E. Peterson, and M. Kiker. "Concept Based Lexical Selection". In *AAAI 1994 Fall Symposium on Knowledge Representation for Natural Language Processing in Implemented Systems*, New Orleans, LA, 1994.

[Dorr, 1993] B. Dorr. *Machine Translation: A View from the Lexicon.* MIT Press, Cambridge, MA, 1993.

[Heine *et al.*, 1991] B. Heine, U. Claudi, and F. Huennemeyer. *Grammaticalization: A Conceptual Framework.* University of Chicago Press, Chicago, IL, 1991.

[Herskovits, 1986] A. Herskovits. *Language and Spatial Cognition.* Cambridge University Press, Cambridge, England, 1986.

[Jackendoff, 1983] R. Jackendoff. *Language and Cognition.* MIT Press, Cambridge, MA, 1983.

[Jackendoff, 1990] R. Jackendoff. *Semantic Structures.* MIT Press, Cambridge, MA, 1990.

[Jackendoff, 1991] R. Jackendoff. "Parts and Boundaries". In B. Levin and S. Pinker, editors, *Lexical and Conceptual Semantics.* Blackwell Publishers, Cambridge, MA, 1991.

[Langacker, 1987] R. Langacker. *Foundations of Cognitive Grammar, Vol. 1: Theoretical Prerequisites.* Stanford University Press, Stanford, CA, 1987.

[Levin, 1993] B. Levin. *English Verb Classes and Alternations: A Preliminary Investigation.* University of Chicago Press, Chicago, IL, 1993.

[Onyshkevych and Nirenburg, 1994] B. Onyshkevych and S. Nirenburg. "The Lexicon in the Scheme of KBMT Things". Technical Report MCCS-94-277, New Mexico State University, 1994.

[Pugeault *et al.*, 1994] F. Pugeault, P. Saint-Dizier, and M. G. Monteil. "Knowledge Extraction From Texts: A Method For Extracting Predicate-Argument Structures From Texts". In *Proceedings of Fifteenth International Conference on Computational Linguistics*, Kyoto, Japan, 1994.

[Retz-Schmidt, 1988] G. Retz-Schmidt. "Various Views on Spatial Prepositions". *AI Magazine*, pages 95–105, Summer 1988.

[Reyero-Sans and Tsujii, 1994] I. Reyero-Sans and J. Tsujii. "A Cognitive Approach to an Interlingua Representation of Spatial Descriptions". In *AAAI Workshop on Integration of Natural Language and Vision Processing*, Seattle, Washington, 1994.

[Talmy, 1985] L. Talmy. "Lexicalization Patterns: Semantic Structure in Lexical Forms". In T. Shopen, editor, *Language Typology and Syntactic Description 3: Grammatical Categories and the Lexicon*, pages 57–149. University Press, Cambridge, England, 1985.

[Verrière, 1994] G. Verrière. "Manuel d'utilisation de la structure lexicale conceptuelle (LCS) pour représenter des phrases en français". Research note, IRIT, Université Paul Sabatier, Toulouse, France, June 1994.

[Voss and Dorr, 1995] C. Voss and B. Dorr. "Toward a Lexicalized Grammar for Interlinguas". *Journal of Machine Translation*, 1995. To appear.

[Wu and Palmer, 1994] Z. Wu and M. Palmer. "Verb Semantics and Lexical Selection". In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pages 133–138, Las Cruces, New Mexico, 1994.