Multilingual Generation: The Role of Telicity in Lexical Choice and Syntactic Realization

Bonnie J. Dorr and Mari Broman Olsen
Institute for Advanced Computer Studies
Department of Computer Science
University of Maryland
College Park, MD 20742

(Received ..... ; Accepted in final form .....)

Abstract. Multilingual generation in machine translation (MT) requires a knowledge organization that facilitates the task of *lexical choice*, i.e., selection of lexical units to be used in the generation of a target-language sentence. This paper investigates the extent to which *lexicalization patterns* involving the *lexical aspect* feature [+telic] ('having an inherent end' as in *win* vs. *run*) may be used for translating events and states among languages. The work reported here is part of an ongoing effort to build a computational model that characterizes knowledge at the syntax-semantics interface [18, 37, 38, 39, 65, 66]. Telicity has been correlated syntactically with both transitivity [65, 66, 69] and unaccusativity [38, 39], and semantically with Talmy's 'path' of a motion event [62, 63, 64], the representation of which characterizes languages parametrically.

Taking as our starting point the syntactic/semantic classification in Levin's *English Verb Classes and Alternations* [37], we examine the relation between telicity and the syntactic contexts, or *alternations* outlined in this work, identifying systematic relations between the lexical aspect features and the semantic components that potentiate these alternations. Representing lexical aspect—particularly telicity—is therefore crucial for the tasks of lexical choice and syntactic realization. Having enriched the data in [37] (by correlating the syntactic alternations (Part I) and semantic verb classes (Part II) and marking them for telicity) we assign to verbs *lexical semantic templates* (LSTs). We then demonstrate that it is possible from these templates to build a large-scale repository for *lexical conceptual structures* (LCSs)—in the spirit of a number of researchers [20, 27, 31, 32, 38, 39]—which encode meaning components that correspond to different values of the telicity feature. The LST framework preserves both *semantic content* and *semantic structure* (following Grimshaw [24]) during the processes of lexical choice and syntactic realization. Application of this model identifies precisely where the Knowledge Representation (KR) component may profitably augment our rules of composition, to identify cases where the IL underlying the SL sentence must be either reduced or modified in order to produce an appropriate TL sentence.

Key words: syntax-semantics interface, syntactic alternations, semantic classes, (a)telicity, multilingual generation, interlingua, lexical conceptual structure

1. Introduction

The problem of multilingual generation in applications such as machine translation (MT) requires a knowledge organization that facilitates the task of *lexical choice*, i.e., selection of lexical units to be used in the generation of a target-language sentence. This paper focuses on the development of a theory of lexical choice—particularly for the task of MT—for verbs and their arguments in multiple languages. We use the syntactic contexts, or *alternations*, and semantic classes in [37] to lay a foundation for exploiting *lexicalization patterns* [62, 64] in translating situations (events or states) among languages.

The work reported here is part of an ongoing effort to build a computational model that characterizes knowledge at the syntax-semantics interface [18, 37, 38, 39, 65, 66], taking into account the requirements of the translation task. We aim to identify the linguistically relevant properties of predicate meaning that may be used in a translation *interlingua* (language-independent representation) and proposed as candidates for linguistic universals.

The *lexical aspect* feature [+telic] (i.e., having an inherent end, as in *win* vs. *run*) plays a key role in the characterization of knowledge at the syntax-semantics interface. It is therefore crucial for the tasks of lexical choice and syntactic realization. From the syntactic point of view, this feature has been correlated with both transitivity [65, 66, 69] and unaccusativity [38, 39]. Moreover, there is increasing evidence that telicity and related notions (e.g., Verkuyl's 'terminative aspect' [73] and Tenny's 'measure' and 'terminative' roles [66] interact "more profoundly with grammar and syntax" than other lexical aspect features, such as durativity ([+durative]) [67, p. 123]. From the semantic point of view, telicity overlaps with Talmy's notion of 'path' [62, 63, 64]: motion events with source or goal paths are [+telic], e.g., *go out/in*.

Languages vary considerably with respect to how telicity is realized in the corresponding surface structures. Languages such as Korean and the Romance languages tend to lexicalize paths in verbs (e.g., Spanish *entrar* 'enter'). They may therefore have (or perhaps employ) more [+telic] verbs (cf. [59]) and possibly more alternations that indicate telicity. In contrast, English, Mandarin, and non-Romance Indo-European languages tend to lexicalize paths external to verbs, as in *go in*. They may therefore have (or perhaps employ) fewer [+telic] verbs, as well as more other constituents carrying telicity, and possibly more alternations that add [+telic] constituents. Furthermore, the typological differences in lexicalization patterns have parallels in a wide range of conceptual fields [64]. Telicity therefore plays a critical role in determining how event structure maps onto surface structure. For translation, then, one might envision an algorithm based upon this typology, such that Spanish [+telic] verbs map onto English verbs and an appropriate [+telic] constituent (see [44, p. 352ff]).

Our goal is to investigate the role of telicity in the characterization of lexical knowledge; we then use the results of this investigation to refine meaning components in our interlingual representation and enhance the lexical choice process. We take as our starting point the syntactic/semantic classification in Levin's *English Verb Classes and Alternations* [37]. We examine the relation between the [+telic] feature and the syntactic contexts, or *alternations*, outlined in this work, identifying alternations that indicate (a)telicity or have a telicizing effect on the relevant verbs. We verify the hypothesis that lexical aspect features need not be primitive but may be derived from the same semantic components that potentiate the alternations [47]. We then use the categorization of the alternations with respect to telicity to assign a telicity value to the verbs in each of Levin's 190 semantic classes, according to their participation in the alternations. Finally, we demonstrate that it is possible to use these semantic components in a large-scale repository of *lexical conceptual structures* (LCSs)—in the spirit of a number of researchers [20, 27, 31, 32, 38, 39]—which encode meaning components that reflect different values of the telicity feature. This repository, which is hierarchically organized, serves as the basis of a lexical-choice module of a MT system [15, 17, 18, 19].

The following section describes the methodology and assumptions of our approach, in particular, the syntax-semantics relation upon which our study is based and the relation of this work to alternative approaches. Section 3 defines a mapping between Levin's 84 syntactic alternations and the [+telic] feature. Section 4 presents the assignment of a telicity value to verbs in each of Levin's 190 semantic classes. In Section 5, we use this telicity value to construct LCSs for interlingual MT. Finally, Section 6 gives an example of how our hierarchical LCS organization, stored in a structure called the *IL Lexicon*, is used for the lexical choice task in a multilingual context.

2. Methodology and Assumptions

The nature of lexical aspect (also known as *Aktionsart*) has been the subject of much debate, in linguistic as well as in philosophical circles. Some have argued that the temporal properties represented by lexical aspect features fall outside the domain of semantics proper. Porter [51], for example, adopts this position because the lexical aspect features—unlike the grammatical aspect features [+imperfective] and [+perfective]—are not morphologically realized. In contrast, Vendler [71, p. 102] demonstrates that lexical aspect is semantically relevant by showing how his four categories (STATE, ACTIVITY, ACCOMPLISHMENT, and ACHIEVEMENT) behave differently in a variety of temporal test frames, such as *For how long did* X V*?* and *How long did it take* X *to* V*?*. Accomplishments, for example, are odd with the first frame

(*#For how long did John build a house?*) but not the second (*How long did it take John to build a house?*) (cf. [4, 20, 22, 42, 53, 72]).[1]

Olsen [45, 46] sides with the Vendler tradition, showing furthermore that these lexical aspect classes are not themselves semantic primitives but composed of three privative lexical aspect features: [+durative], [+dynamic], and [+telic].[2] Verbs marked [+durative] denote situations that persist over a temporal interval; those unmarked for durativity ([∅durative]) are usually interpreted as punctiliar, although they may be interpreted as durative, given the appropriate constituents or pragmatic context. Verbs marked [+dynamic] are events; those unmarked for dynamicity ([∅dynamic]) are usually interpreted as states. The [+telic] verbs are interpreted as having an inherent end and [∅telic] verbs as generally lacking such an end. It is therefore lexical aspect features that condition different semantic behavior in the syntactic test frames. Where the features are unmarked for a given class, verbs in the class show variable behavior as well. For example, "activity" verbs (e.g., *run*), since they lack the feature [+telic], may be interpreted either as activities (*John swept the driveway*) or as accomplishments (*John swept the leaves off the driveway*), depending on other constituents and the pragmatic context.

This model of lexical aspect contrasts with models that treat the classes either as indivisible primitives or as made up of equipollent features: [±dynamic] (dynamic/stative), [±durative] (durative/punctiliar), and [±telic] (telic/atelic).[3] In the privative model, verbs are minimally either [+dynamic] or [+durative]. Other features may be marked or left unmarked, as summarized in (1) [45, pp. 32–33] (cf. charts in [7, p. 57], [60]).

(1)

Aspectual Class	**Telic**	**Dynamic**	**Durative**	**Examples**
State			+	*know, have*
Activity		+	+	*run, paint*
Accomplishment	+	+	+	*destroy*
Achievement	+	+		*notice, win*
Semelfactives		+		*cough, tap*
Stage-level states	+		+	*be sick*

[1] The symbol "#" marks semantic anomaly.

[2] The features and a cooccurrence restriction permit description of two other categories: semelfactives and stage-level states [45, p. 66].

[3] The equipollent analysis is the one assumed in earlier work [10], where *ad hoc* mechanisms (e.g., coercion functions) were used to account for phenomena that are accounted for more systematically by the privative analysis.

The marked lexical aspect features are always present in the semantic representation, but features that are unmarked on the verb may become marked by other constituents.[4] The model therefore accounts for variance in lexical aspect interpretation by means of monotonic composition.

Consider the *Wipe* Alternation [37, p. 53], where an activity verb (*sweep*) is used to represent both an activity (marked [+dynamic, +durative]) and an accomplishment (marked [+dynamic, +durative, +telic]):

(2) **2.3.3 Wipe Alternation:**
A. John swept the driveway. ACTIVITY
B. John swept the leaves off the driveway. ACCOMPLISHMENT

Note that an *alternation* consists of (A) a *basic* variant and (B) a *non-basic* variant.[5] The privative model allows one to describe not only which verbs have a given inherent feature (i.e., in the A variant), but it also assumes that other elements added in syntactic alternations—such as the direct object in the B variant of (2)—may mark lexical aspect features. Wherever appropriate (and except as noted) we refer to the alternations by the name of the non-basic variant. For example, when we refer to the *Wipe* Alternation as adding telicity, we mean the variant with the direct object and locative prepositional phrase (the B variant), rather than the one with just a direct object (the A variant).

Vendler distinguishes the temporal properties of verbs (represented here by lexical aspect features) from their non-temporal semantics. It remains to be seen, however, whether lexical aspect is indeed orthogonal to other lexical semantic properties. Since differences in lexical aspect condition different behavior in syntactic frames, lexical aspect features—in particular, telicity—appear to be among the semantic properties on which the alternations catalogued by Levin [37] depend. We therefore propose a systematic relation between lexical aspect—in particular, telicity—and these "linguistically ... pertinent aspects of verb meaning" [37, p. 1]. As in [16], aspect is taken to be a property of interlingual structures of type SITUATION. Our approach

[4] They may also be interpreted as marked in the appropriate pragmatic context [45, p. 57].

[5] We label as non-basic the construction that the alternation names. However, seven of Levin's alternations are associated with only one variant corresponding to the named syntactic construction (e.g., the Obligatory Passive alternation). In addition, 18 alternations have two variants named by the alternation, neither of which is more basic than the other (e.g., the With/Against alternation). Without loss of generality (formally speaking), we assume that an alternation consists of an A/B pair: For the singleton cases, we *add* an applicable basic form, and for the cases where neither variant is basic, we we *split* the alternation into two and then *add* a basic form to each one.

differs from this work, however, in that we do not consider lexical aspect features to be primitive and independent, but related to—and perhaps derived from—other semantic elements that are represented in the interlingua, particularly change of state.

Levin [37] hypothesizes that semantically-related verbs share certain syntactic alternations. For example, verbs in the *Break* semantic class (*break, crack, rip, shatter, snap, ...*) participate in the Middle alternation,[6] but not the Conative or Body-Part Possessor Ascension alternations:[7]

(3) (i) **1.1.1 Middle:**
A. She broke the window.
B. The window breaks easily.

(ii) **1.3 Conative:**
A. She cracked the ice.
B. ∗She cracked at the ice.[8]

(iii) **2.12 Body-Part Possessor Ascension:**
A. She snapped John's collar bone.
B. ∗She snapped John on the collar bone.

Verbs in the *Hit* semantic class (*bash, hit, kick, pound, tap, whack, ...*) differ from those in the *Break* semantic class in that they participate in the Conative and Body-Part Possessor Ascension alternations, but not in the Middle alternation:

(4) (i) **1.1.1 Middle:**
A. She hit the window.
B. ∗The window hits easily.

(ii) **1.3 Conative:**
A. She tapped the ice.
B. She tapped at the ice.

(iii) **2.12 Body-Part Possessor Ascension:**
A. She whacked John's collar bone.
B. She whacked John on the collar bone.

A more detailed investigation reveals that four semantic verb classes are uniquely identified according to their participation (or lack thereof) in the above three alternations (after [37, p. 7]).

[6] Although English does not mark the middle voice, the construction in (3)(i)B parallels the "facilitative middle" in languages with such a marking [33, p. 268].

[7] Section numbers (and, in several cases, example sentences) are taken from [37]. Numbers less than 10 refer to alternations; those greater than or equal to 10 refer to verb classes.

[8] Syntactically ill-formed sentences are marked with an asterisk (∗).

(5)

Verb Class	1.1.1	1.3	2.12
Break: break, crack, rip, shatter, snap, ...	yes	no	no
Hit: bash, hit, kick, pound, tap, whack, ...	no	yes	yes
Touch: pat, stroke, tickle, touch, ...	no	no	yes
Cut: cut, hack, saw, scratch, slash, ...	yes	yes	yes

In fact, Levin surveys 190 semantic classes, each of which is characterized by a subset of 84 possible syntactic alternations (on the average, 4 alternations per class).

Levin's central thesis—that the semantics of a verb and its syntactic behavior are predictably related—is tested and verified in [13]. Building on this work, Dorr and Jones [14] develop techniques for automatic extraction of semantic classes for 7757 English verbs (approximately 5000 of which are not given in Levin's classes) using purely syntactic alternation tests such as (3) and (4) above. Although this work addresses the problem of defining a mapping between semantics and syntax, it does not define the *meaning components* underlying the semantic portion of the mapping. That is, although we know what semantic class a verb belongs to (e.g., using the automatic techniques in [13]), the class does not provide a semantic representation for that verb. As Levin observes [37, p. 15], the task of deciding "what components of verb meaning figure in the relevant generalizations" about syntactic behavior may be undertaken only after one has established "to what extent the meaning of a verb determines its syntactic behavior."

Our goal is to identify the underlying meaning components for each verb class; these components will then serve as the basis of the development of an *interlingual* representation for lexical choice in multilingual generation. We found telicity to be a key distinguishing feature in the characterization of verb meaning in many classes. We have therefore investigated the semantics surrounding the 84 alternations in [37], which apply on the basis of verb class membership. We have examined the semantic elements that potentiate the alternations, as well as the semantics of the alternations themselves, and categorized the alternations into five classes with respect to telicity: (i) alternations that indicate telicity (all participating verbs are inherently [+telic] in the basic (A) variant); (ii) alternations that add telicity (all participating verbs are [+telic] in the non-basic (B) variant); (iii) alternations that indicate atelicity (all participating verbs are inherently [$\emptyset$telic], i.e., in the basic (A) variant); (iv) alternations and constructions that are irrelevant with respect to (a)telicity (some participating verbs are inherently [+telic] and others inherently [$\emptyset$telic], and their categorization is not systematically affected by the relevant construction); and, for completeness, (v) a small number of alternations that cannot be classified. We use the above categorization to enrich the data in [37], correlating the syntactic alternations (Part I) and semantic verb classes (Part II) and marking them for telicity. We then use this as the basis of assigning to verbs lexical representations which then serve as a foundation for large-scale lexicons.

For the purpose of lexical choice, the telic/atelic distinction is important in that it correlates with a difference in the realization of constituents in the output sentence. As an example, consider the *removal* category. Within this category, the behavior of verbs in a [+telic] class, such as *clean,* contrasts syntactically with verbs in a [∅telic] class, such as *sweep*. Only in the case of *sweep* is a telicizing alternation allowed, such as the change-of-state resultative construction:[9]

(6) (i) * John cleaned the driveway sparkling.

(ii) John swept the driveway clean.

On the other hand, an *of* phrase may be used for *clean*, but not *sweep*:

(7) (i) John cleaned the driveway of leaves.

(ii) * John swept the driveway of leaves.

We assume that verbs, or more accurately verb senses, that share a semantic class have the same basic lexical structure and the same set of constraints on possible syntactic-realization operations. As an example of a syntactic-realization operation consider the resultative construction in (6) above: this operation is applicable if the chosen target-language verb is [∅telic]—e.g., *sweep* (in the *Wipe* verb class)—but not if the chosen target-language verb is [+telic], *clean* (in the *Clear* verb class). We will address such examples further in Section 6.

Our approach is similar to that of [9], which also uses verb classification system for organizing lexical entries, but with the complementary goal of handling fine-grained, stylistic variations. There are also similarities between our approach and that of [43], in that we seek a single unifying data structure to establish a range of semantic relations among words.[10] Because our primary focus is on the development of

[9] Occasionally it is possible for a result verb to participate in a resultative construction, but only if the result phrase further specifies the verb's inherent result as in *I drained the tub dry*. In such "non-canonical" examples, we would expect human judgments to vary with respect to grammaticality.

[10] Strictly speaking, the lexicons in interlingua-based MT systems are not restricted to word-level entries. For the purposes of this paper however we will refer to "words" in the lexicons, setting aside the details about other types of lexical entries. See [40] for further discussion on extending the range of lexical entries in MT systems.

the interlingua and associated lexicons, our lexical entries are driven by our lexicalization needs and the primitives in the interlingual syntax. (By contrast, in many large-scale, knowledge-based MT projects, the concepts in the knowledge base drive the definitions of lexical entries.)

Alternative lexical choice frameworks have been explored by a number of researchers [1, 30, 48, 55, 57, 61]. In the approach of Palmer and Wu [48], lexical mismatches are resolved through the use of a conceptual lattice based on an adapted version of the classification in [37]. While this approach is similar in spirit to our method of mismatch resolution, Palmer and Wu propose an *approximate* matching scheme based on "semantic closeness" as determined by numerical distances between semantic classes, whereas our scheme relies on a full *structural-coverage* matching scheme, i.e., alignment among lexical-semantic components that systematically map onto syntactic structure. We argue in Section 5.1, that we must preserve substructure information (i.e., LCS descriptions) in order to provide an adequate translation into a target-language surface structure; thus, a numerical measure of semantic closeness is not adequate for our purposes. Moreover, the structural-coverage matching scheme is a prerequisite for recursive instantiation of substructures that are not necessarily aligned with the full structure of the interlingua during the translation process. Finally, Levin's classification system [37] has been refined in later work [38, 39], where regularities in lexical representations—and the rules that operate on these representations—can be exploited for a more economical approach to MT. Such regularities would be lost if we were to adopt a numerical "closest match" scheme.

The novelty of our approach is that, in the spirit of work by a number of researchers [24, 38, 50, 63], the interlingual representation is designed to preserve the linguisticaly relevant *semantic structure* of the source- and target-language sentences while still retaining the idiosyncratic *semantic content* underlying the interlingual primitives. The preservation of semantic structure, representing meaning shared by members of verb classes, allows us to readily map between the interlingua and the syntax (syntactic realization); the preservation of the constants in the semantic content allows us to readily map between the IL and target-language words (lexical choice). The advantage to this approach is two-fold: (i) the lexical choice task is constrained according to constants in the basic lexical representations, which are systematically related to surface lexical items; (ii) the syntactic realization task is constrained according to the argument-taking and telicity properties of the chosen target-language verb. Approaches that ignore the structural nature of lexical items and their combination stand to lose the benefit of regularities that exist *within* a language (in the lexicon) as well *across* languages (in the surface syntactic form). By decomposing verbs into predicate *structure* and non-predicate *constants*, we exploit such regularities.

3. Relation Between Telicity and Syntactic Alternations

This section provides an analysis of verbs and predicates from the alternations and constructions in Part I of [37]. In categorizing the alternations, it was first determined whether the verbs undergoing the alternation were a homogeneous class with respect to telicity. The assignment of alternations to categories was based on testing the verbs and classes as positive exemplars of these [37, p. 19]. Levin illustrates each alternation with representative semantic classes, with indications of acceptability based primarily on intuition (Levin, p.c.). We supplemented these with our own intuitions, adding or eliminating classes or verbs, as noted below. Appendix A summarizes the classification of alternations into the five telicity categories described below.

3.1. Alternations Indicating [+telic]

The first telicity category identifies [+telic] verbs, i.e., verbs undergoing the alternation are all [+telic] on standard tests, in particular, the entailment test shown in (8) (see, e.g., [22, p. 109], [34, p. 172], [21, p. 46], and [20, p. 57]).[11]

(8) (i) John was running ENTAILS John has run. ATELIC

(ii) John was winning DOES NOT ENTAIL John has won. TELIC

As an example of such a case, consider the verbs undergoing the intransitive version of the Total Transformation Alternation:

(9) **2.4.4 Total Transformation Alternation (intransitive):**
A. He turned into a frog.
B. He turned from a prince into a frog.
Verbs: TURN VERBS (some): alter, change, metamorphose, ?transform, ?transmute, turn

[11] The test is simplified here, since the privative analysis permits atelic verbs to have telic interpretations and the corresponding entailments. More accurately, we consider verbs with uniform lack of entailments as telic and those with variable behavior as atelic (see [45, p. 12] for discussion).

All verbs undergoing this alternation (a subset of the TURN VERBS listed by Levin [37]) are [+telic] according to standard tests such as the following:[12]

(10) John was turning DOES NOT ENTAIL John has turned.

3.2. Alternations Adding [+telic]

The set of alternations adding [+telic] consists of those alternations that applied either to a uniformly atelic set of verbs or to a mixed ([+telic] and atelic) set of verbs. These alternations are classified as **adding** [+telic]. An example of such a case is the Cognate Object Construction as shown in (11) [37, pp. 95–96], [41].

(11) **7.1 Cognate Object Construction:**
A. I fought.
B. I fought the good fight.
Verbs: die, live, sing, smile, tell...

Several additional alternations are classified as adding [+telic]. Of these, 10 apply uniformly to atelic predicates and may therefore be used as [∅telic] diagnostics. The Conative, for example, applies only to atelic verbs like *hit* and not to atelic verbs like *break*:

(12) Frank Thomas hit at the low slider
∗The ball broke at the scoreboard

3.3. Alternations Indicating [∅telic]

The next telicity category identified the set of alternations that indicate [∅telic]. In these cases, both the bare verbs and the verbs in the non-basic variant are uniformly atelic. As an example, consider the *With/Against* Alternation, with the verbs from [37] exhaustively listed:

(13) **2.8 With/Against Alternation:**
A. I hit the fence with the rod.
B. I hit the rod against the fence.
Verbs: HIT VERBS: bang, bash, batter, beat, bump, butt, dash, drum, hammer, hit, kick, knock, lash, pound, rap, slap, smack, smash, strike, tamp, tap, thump, thwack, whack

[12] Constructions that indicate telicity of verbs to which they apply are also themselves [+telic], since marked features may not be removed under the monotonic composition of privative features introduced in [45, 46]. That is, if the verbs themselves are all [+telic], so are both variants of an alternation.

3.4. Alternations for which (A)telicity is Irrelevant

If neither the verbs nor the alternation variants are homogeneous [+telic] or [∅telic] classes, then the alternation is classified as irrelevant with respect to telicity, as illustrated by the *With* Preposition Drop Alternation.

(14) **1.4.2 With Preposition Drop:**
A. Jill met (with) Sarah.
B. Jill battled (with) Sarah.
Verbs: MEET VERBS: battle, box, consult, debate, fight, meet, play, visit

3.5. Unclassifiable Alternations

Two alternations could not be classified, since Levin [37] gives only a handful of verbs:

(15) **Unclassifiable Alternations:**
3.2 Natural Force Subject Alternation: dry
3.3 Instrument Subject Alternation: break, eat, load

3.6. Summary of Categorized Alternations

The categorization of the 84 alternations may be summarized as follows:

(16)

Telicity Property	
Indicates [+telic]:	7
Adds [+telic]:	27
Indicates [∅telic]:	13
Irrelevant to (a)telicity:	37
Unclassifiable:	2
TOTAL:	84

4. Relation Between Telicity and Semantic Classes

Once we categorized all of Levin's alternations into telicity types, we assigned telicity values to the semantic classes according to the telicity types associated with the relevant alternations. As part of this task, we produced a database containing pairings between alternations in Part I and semantic classes in Part II, using as a starting point the publicly available online index from Levin's *English Verb Classes and Alternations* [37].[13] Since the verbs and classes in Part I are only illustrative of the properties of the relevant alternations, they are not an exhaustive list of the verbs and verb classes participating in the alternations. Similarly, Part II lists only "central properties" [37, p. 19] of the verb classes, and does not give all alternations in which a verb class participates, or from which it is prohibited. Therefore, we used a combination of techniques—semi-automatic compilation and human verification—to convert the online index into a database where Parts I and II of the book were consistently cross-indexed, and enhanced with respect to grammaticality judgements.[14]

In order to construct this cross-indexed database, we first enhanced the original online index in a number of ways:

- **Non-alternating verbs:**
  The verb *decouple* is paired with the *Apart* Alternation, 2.5.6, in the online index:

  ```
  decouple ...2.5.6...
  ```

  However, this index listing refers to a case where the alternation is *not* applicable. In fact, the alternation does not apply to any of the TURN VERBS, 23.1. We have enhanced the online index to capture this fact by means of a class/alternation pairing combined with the $*$ notation to indicate ungrammaticality:

  ```
  decouple ...(23.1 *2.5.6)...
  ```

- **Polysemous verbs:**
  The verb *relieve* is paired with the *Clear* Alternation, 2.3.2, in Levin's online index:

  ```
  relieve ...2.3.2...
  ```

[13] The file evca93.index is located in the directory linguistics/texts/indices in the Michigan linguistics archives. The Internet address is linguistics.archive.umich.edu. The file itself is 89411 bytes in size and is an ASCII (text) file that may be downloaded in text mode by anonymous ftp.

[14] We are indebted to Joseph Rosenzweig at the University of Pennsylvania—who simultaneously undertook a similar task—for his useful guidance and commentary during creation of the cross-index between Parts I and II.

Since *relieve* is in two classes (10.6 POSSESSIONAL DEPRIVATION VERBS and 31.1 AMUSE VERBS), a raw compilation of Levin's original index would result in both of these classes being paired with alternation 2.3.2. However, this alternation applies only to the first of the two classes (10.6). The incorrect pairing of 2.3.2 with 31.1 is symptomatic of the *polysemy* problem—discussed by a number of researchers [3, 8, 29, 35, 54, 68, 75]—which, for our study, corresponds to the occurrence of verbs in multiple semantic classes and alternations.[15] We have resolved this by enhancing the online index so that the two semantic classes are distinguished with respect to alternation 2.3.2:

```
relieve ...(10.6 2.3.2) (31.1 *2.3.2)...
```

– **Multiple variants:**
The verb *saddle* is paired with the *Fulfilling* Alternation, 2.6, in Levin's online index:

```
saddle ...2.6...
```

However, in this case, only one of the two variants (the B variant) is applicable. So we enhanced the online index by adding a 'B' prefix to the relevant number:

```
saddle ...(13.4.2 B2.6)...
```

There are also cases where only the A variant applies, but these typically were not listed in the original online index. Thus, we added the A variant cases through human inspection, e.g., the verb *lend* in class 13.1 participates in the basic 'A' variant of the *Fulfilling* Alternation:

```
lend ...(13.1 A2.6)...
```

In summary, we have enhanced the online index by pairing the semantic classes with the applicable/nonapplicable alternations and annotating them as follows: A (where only alternate A is allowed); B (where only alternate B is allowed); + (where both alternates are allowed); and $*$ (where neither alternate is allowed).

From this enhanced index, we generated the cross index between parts I and II, as illustrated by the entry for HIT VERBS:

[15] Ullman [68, p. 161], for example, finds five transitive and five intransitive senses of the verb *rush* in the *Shorter Oxford Dictionary*. See [13, 14] for a more detailed discussion of the effect of polysemy in lexical acquisition using data from [37].

(17) 18.1 HIT VERBS:
∗2.1 Dative
+1.2.6.1 Characteristic Property of Agent
+1.2.6.2 Characteristic Property of Instrument
+1.3 Conative
+2.12 Body-Part Possessor Ascension
+2.5.2 Together Reciprocal (trans)
+2.8 With/Against
+7.5 Resultative
+7.6.1 Unintentional Interpretation of Reflexive Object
+7.6.2 Unintentional Interpretation of Body-Part Object
A1.1.1 Middle
A1.1.2.1 Causative/Inchoative
A1.1.2.3 Other Causative/Inchoative
A2.5.1 Simple Reciprocal (trans)
B2.9 Through/With

Once the cross-index was created, we then undertook the task of telicity assignment to each semantic class. There were two phases to this task. In the first phase, we categorized semantic classes into telicity types for those cases where the relevant alternations were uniformly either [+telic] or [∅telic] indicators. (At this stage we ignored irrelevant and unclassifiable alternations.) Out of the 190 semantic classes, 86 fell into one of these two categories listed in Appendix B:

(18) **Number of Telic and Atelic Semantic Classes:**
Classes associated with [+telic] indicators: 15
Classes associated with [∅telic] indicators/diagnostics: 71

In the second phase of the telicity assignment, we divided the remaining 104 semantic classes into two categories: those that were associated with conflicting telicity indicators/diagnostics and those that were not associated with telicity indicators/diagnostics (i.e., unknown with respect to telicity):

(19) **Number of Unidentifiable Semantic Classes:**
Classes associated with conflicting indicators/diagnostics: 3
Classes not associated with telicity indicators/diagnostics: 101

The three conflict cases are given below:

(20) (i) 22.2 AMALGAMATE VERBS (*affiliate, alternate, amalgamate*)
3.7 Container Subject [atelic]
2.5.4 Simple Reciprocal (intrans) [telic]

(ii) 37.1 VERBS OF TRANSFER OF A MESSAGE (*ask, cite, demonstrate*)
1.2.1 Unspecified Object [atelic]
3.4 Abstract Cause Subject [telic]

(iii) 37.5 TALK VERBS (*speak, talk*)
1.2.7 *Way* Object [atelic]
7.4 *X's Way* [atelic]
2.5.4 Simple Reciprocal (intrans) [telic]

The conflict in each of the three cases may be attributed to certain fine-grained distinctions made within individual verb classes in [37]. For example, class 37.1—VERBS OF TRANSFER OF A MESSAGE—is mixed with respect to participation in the two relevant alternations: Unspecified Object makes atelic predicates telic and applies to *write*, but not to *explain*); Abstract Cause Subject applies only to telic predicates, i.e. to *explain*, but not *write*). In fact, these two alternations apply to verbs in a complementary distribution, i.e., there are no conflicts for individual verbs.[16]

Out of the 101 "unknown" verb classes, 45 are associated with 13 alternations relevant to telicity. We found it possible to provide additional categorization of semantic classes by assigning a "reverse" telicity value to the negative versions of the relevant alternations. We considered an alternation to be a *negative version* if one or more non-basic variant(s) were ruled out for a particular class. For example, the negative version of the *Conative* Alternation is associated with the DESTROY VERBS, class 44 (*annihilate, blitz, decimate, demolish, destroy*):

(21) **1.3 Conative Alternation:**
A. The storm destroyed the ship
B. *The storm destroyed at the ship

Since the *Conative* Alternation is an atelicity indicator, we take all classes associated with the negative version to be [+telic].[17]

[16] It might be possible that another aspectual feature is relevant. For example, Olsen [47] has investigated alternations indicating [∅dynamic]; the verbs in class 22.2 are uniformly [∅dynamic]. In addition, some of the "unknown" verb classes are uniformly [∅dynamic], e.g., class 54.3 FIT VERBS.

[17] For this phase of the study, we eliminated from consideration the negative version of the following two telicity indicators—Total Transformation alternation (transitive) (2.4.3) and Fulfilling alternation (2.6)—both of which have two non-basic variants. These were eliminated when it was determined, through human inspection, that verbs that do not participate in these alternations are often telic:

(i) **Total Transformation:**
*The Romans destroyed the city into a ruin
*The Romans destroyed the city from a capital into a ruin

The decision to provide additional semantic classification using the reverse assignment scheme was based on the observation that at least some of these alternations are incompatible with the notion of [+telic]. That is, if an alternation adds telicity to uniformly [∅telic] verbs (e.g., the *Conative* alternation), perhaps verbs that do not participate in this alternation are uniformly [+telic]. Such alternations may therefore be used to identify [∅telic] verbs and predicates.[18] Using this reverse assignment scheme, the 13 additional alternations are categorized as follows:

(22) (i) **Negative Versions of Alternations Indicating [∅telic]:**
2.5.4 Simple Reciprocal Alternation (intransitive)

(ii) **Negative Versions of Alternations Indicating [+telic]:**
1.1.3 Substance/Source Alternation
1.2.1 Unspecified Object Alternation
1.3 Conative Alternation
2.3.1 *Spray/Load* Alternation
2.3.3 *Wipe* Alternation
2.10 *Blame* Alternation
2.11.3 STALK VERBS: NP through/in NP
2.11.4 RUMMAGE VERBS: through/in NP for NP
2.11.5 FERRET VERBS: NP out of NP
2.13.3 Possessor and Attribute Object Alternation
3.8 Raw Material Subject Alternation
7.5 Resultative Construction

With these hypothesized assignments, the 45 "unknown" classes were categorized into telic and atelic categories as shown in Appendix C. Manual inspection of these classes confirms the assignment of 31 of 45 classes on the relevant interpretations, with exceptions annotated with the '-' symbol. Those marked 'm' are mixed in telicity: these alternations were classified based on an alternation that applies to a uniformly (a)telic subset.[19]

(ii) **Fulfilling:**
He offered a job to her
*We offered her with a job

The above sentences contain verbs that are telic by standard tests of telicity; thus, the negative versions of these two alternations were removed from consideration for categorization of verb classes 13.1, 13.2, and 44.

[18] Investigating why these alternations prohibit the [+telic] feature marking may yield additional insights into the nature of telicity, although it is beyond the scope of this study.

[19] All of the semantic classes whose telicity value was assigned on the basis of the Conative Alternation were accurately categorized.

5. Telicity-Based Assignment of LCS Representations to Semantic Classes

The telicity-based categorization of semantic classes above allowed us to assign semantic representations to individual verbs in the identifiable classes. This section defines the representations and outlines the semantic basis for the components thereof.

5.1. Definition of the LCS Representation

We use *lexical conceptual structures* (LCS) [20, 27]—an augmented form of [31, 32]—as the basis of our interlingua.[20] The LCS framework consists of primitives (GO, BE, STAY, etc.), types (Event, State, Path, etc.) and fields (Loc(ational), Temp(oral), Poss(essional), Ident(ificational), Perc(eptual), etc.).

As an example of how the LCS primitive GO is used in representing the sentence semantics of a spatial expression (i.e., in the Loc(ational) field), consider the following case:[21]

(23) (i) The ball rolled toward Beth.

(ii) $[_{\text{Event}}\ \text{GO}_{\text{Loc}}$
$([_{\text{Thing}}\ \text{BALL}],$
$[_{\text{Path}}\ \text{TOWARD}_{\text{Loc}}$
$([_{\text{Thing}}\ \text{BALL}], [_{\text{Position}}\ \text{AT}_{\text{Loc}}\ ([_{\text{Thing}}\ \text{BALL}], [_{\text{Thing}}\ \text{BETH}])])])]$

Roughly, this representation means "The ball changed to a position closer to Beth." Primitive labels (enclosed in square brackets "[]") can take zero or more arguments (enclosed in parentheses "()"). In the lexical entry for the sense of *roll* in (23), the primitive GO takes two arguments: the first is a thematic (affected) object and the second is either a directional path or a means/manner by-phrase.

We adopt a refinement of the LCS representation, incorporating meaning components from the linguistically motivated notion of *lexical semantic template* (LST)—as defined in the work of Levin and Rappaport Hovav [38, 39]. The LST framework provides a decomposition of verbs into structural representations containing *constants*, where a verb with many meanings reflects multiple pairings between constants and structures. Thus, we modify the representation above, recasting it as a causative with two subevents.

[20] See [11] for details of a MT system whose IL is LCS-derived. Recently the LCS framework has been used by others for French, e.g., [52, 74].

[21] The multiple occurrences of the primitive BALL serves only to satisfy requirements on the number of arguments for the primitives GO, TOWARD, and AT; for our sentence generation purposes, these are viewed as a single constituent.

(24) (i) The ball rolled toward Beth.

(ii) $[_{\text{Event}}$ CAUSE
$([_{\text{Event}}$ GO_{Loc} (⟨BALL⟩, [BY ⟨ROLL⟩])],
$[_{\text{Event}}$ GO_{Loc} (⟨BALL⟩,
$[_{\text{Path}}$ $\langle\text{TOWARD}\rangle_{\text{Loc}}$
$([_{\text{Thing}}$ ⟨BALL⟩],
$[_{\text{Position}}$ AT_{Loc} $([_{\text{Thing}}$ ⟨BALL⟩], $[_{\text{Thing}}$ ⟨BETH⟩])])])])]

Roughly, the representation above means "Something caused the ball to change to a position closer to Beth, moving in a certain manner (rolling)." Here, each open-class item, introduces a constant, denoted with angle brackets. For example, the manner-of-motion constant ⟨ROLL⟩ is introduced by the verb *roll*. The constant corresponds to the linguistically atomic *semantic content*—the idiosyncratic part of the meaning that distinguishes among members of a verb class. Two other open-class words, *ball* and *Beth*, are also represented by constants, ⟨BALL⟩ and ⟨BETH⟩. The remainder of the representation is the linguistically analyzed *semantic structure*, as described in [24, p. 1], following a number of researchers including [50, 63]. The systematic use of the LST-style constant in our representations allows us to transcend Jackendoff's framework by providing an entry point into a deeper *knowledge representation* (KR) level, where information such as spatial orientation and temporal duration are provided.

Jackendoff makes the claim that the conceptual structures generalize across fields. In particular, he adopts a *localist* view, claiming that the formalism for encoding constituents in the spatial field generalizes to other fields, at some level of abstraction. Thus, the structure of the representation in (24) carries over to the Identificational field for sentences such as *I swept the driveway clean*, where the ⟨ROLL⟩ constant is replaced by ⟨SWEEP⟩ and the second subevent is a GO_{Ident} instead of a GO_{Loc}. A localist approach to lexical semantics is employed by many others as well [2, 26, 28, 36, 56].

5.2. Semantic Basis for Components of the LCS

In order to assign LCS representations to the individual verbs in Levin's semantic classes, we first examine the semantic components that potentiate the alternations described in Section 3, assuming that the necessary and sufficient semantic conditions may underlie telicity as well. We propose semantic bases for each category, with particular attention paid to alternations indicating and adding telicity. For alternations indicating telicity—category (i)—we examine the semantic components said to potentiate the alternations, and for those adding telicity—category (ii)—the semantic components added by the construction along with telicity.

The results suggest a composite semantic basis for telicity, related to the notion of change of state [20, p. 141], [53, p. 56], but not perfectly correlated with it. Of the 8 alternations **indicating** [+telic] (Section 3.1), 6 of these apply to classes of verbs organized around the property of change of state:

(25) (i) **1.2.3 Understood Reflexive Object Alternation**
A. Jill dressed herself.
B. Jill dressed.
Verbs: DRESS VERBS: bathe, change, disrobe, dress, exercise, preen, primp, shave, shower, strip, undress, wash; LOAD VERBS (some): jam, cram, load, pack; PUSH/PULL VERBS (some): jerk, pull, yank

(ii) **2.3.5 Clear Alternation (intransitive)**
A. Clouds cleared from the sky.
B. The sky cleared (of clouds).
Verbs: CLEAR VERBS (except *clean):* clear, drain, empty

(iii) **2.4.3 Total Transformation Alternation (transitive)**
A. The witch turned him into a frog.
B. The witch turned him from a prince into a frog.
Verbs: TURN VERBS: alter, change, convert, metamorphose, transform, transmute, turn

(iv) **2.4.4 Total Transformation Alternation (intransitive)**
A. He turned into a frog.
B. He turned from a prince into a frog.
Verbs: TURN VERBS (some): alter, change, metamorphose, ?transform, ?transmute, turn

(v) **2.6 Fulfilling Alternation**
A. The judge presented a prize to the winner.
B. The judge presented the winner with a prize.
Verbs: VERBS OF FULFILLING: credit, entrust, furnish, issue, leave, present, provide, serve, supply, trust

(vi) **3.4 Abstract Cause Subject Alternation**
A. He established his innocence with the letter.
B. The letter established his innocence.
Verbs: assert, confirm, demonstrate, establish, explain, imply, indicate, justify, nullify, obscure, proclaim, predict, prove, reveal, show, suggest

All of the above alternations describe processes of transformation or change of either the direct object (in the transitive forms) or the surface subject (in the intransitive forms). The Understood Reflexive Object Alternation in (25)(i) predicates change in, for example, a body, to the state of *clean, dressed,* etc.[22] The *Clear* Alternation in (25)(ii) applies to a subset of "change of state" verbs [37, p. 124]. The Total Transformation Alternations in (25)(iii–iv) describe complete transformations, for which the final state is obligatorily expressed [37, p. 178]. Verbs in the Fulfilling Alternation describe events of transfer—either concrete, as in the example in (25)(iv), or abstract, as in (26):

(26) The supervisor credited the sale to the salesman who closed the deal.

As changes of state, the verb classes to which these apply may be represented as GO_{Ident} or GO_{Loc} situations containing a path, that is, changes of properties or position in the LCS representation.[23]

The semantic features added by 23 of the 27 alternations that **add** [+telic] (Section 3.2) echo the change of state property. According to Levin's description [37], eight alternations indicate change of state or position, two an "affected object", four a "holistic effect" (object completely affected), four "resultatives", and four coming into existence or appearance. All of these may be characterized as changes of state [47]. Some examples are given here:

(27) (i) **Change of State/Position:**
Example: 2.5.3 Apart Reciprocal Alternation (transitive)
A. I broke the twig *off (of)* the branch.
B. broke the twig and the branch *apart*.
Verbs: SPLIT VERBS: blow, break, cut, draw, hack, hew, kick, knock, pry, pull, push, rip, roll, saw, shove, slip, split, tear, tug, yank

(ii) **Affected Object:**
Example: 1.1.1 Middle Alternation
A. Bill pounded the metal flat.
B. This metal won't pound flat.
Verbs: cut, break, ...

(iii) **Holistic Effect:**
Example: 1.4.1 Locative Preposition Drop Alternation
A. Martha climbed up the mountain.
B. Martha climbed the mountain.
Verbs: RUN VERBS (some): canter, climb, cross, fly, gallop,

[22] The PUSH/PULL VERBS predicate change of position, a type of change of state.
[23] Certain of these verb classes may also allow a CAUSE component of meaning. We will see such cases shortly.

hike, jog, jump, leap, prowl, ramble, ride, roam, rove, row, run, shoot (rapids), stroll, swim, traipse, tramp, travel, trudge, vault, wade, walk, wander; VERBS THAT ARE VEHICLE NAMES (some): bicycle, bike, canoe, jeep, raft, row, sail, skate, ski

(iv) **Resultatives:**
Example: 7.5 Resultative Construction
A. The silversmith pounded the metal.
B. The silversmith pounded the metal flat.
Verbs: pound, hammer, push, drink, burn, walk, cry, freeze, slide...

(v) **Coming into Existence/Appearance:**
Example: 4.2 Reflexive of Appearance Alternation
A. I presented a solution to the problem yesterday.
B. A solution to the problem presented itself yesterday.
Verbs: REFLEXIVE VERBS OF APPEARANCE: assert, declare, define, express, form, manifest, offer, pose, present, proffer, recommend, shape, show, suggest

As further support for a privative analysis of telicity, the alternations that indicate atelicity—as well as those irrelevant to (a)telicity (Section 3.4)—fail to show the relatively uniform semantic subclasses of the alternations indicating and adding telicity, i.e., alternations in this category do not cluster around semantic properties. Some (the Conative and the *With/Against* Alternations) entail that their objects undergo no change of state. However, others (*Blame,* Container Subject, and Raw Material Subject alternations) have change-of-state characteristics. The seemingly idiosyncratic behavior of atelic forms, including overlap with characteristics of telic and telicizing alternations, is expected under a privative analysis of telicity, in which only members of the marked class ([+telic]) are semantically homogeneous. The atelic verbs may be represented as non-path $\mathrm{ACT_{Loc}}$, $\mathrm{ACT_{Perc}}$, $\mathrm{GO_{Loc}}$, and $\mathrm{GO_{Perc}}$ situations in the LCS representation.

5.3. Assignment of LCSs to Telic and Atelic Verbs

The distinction between telic and atelic verbs, and their associated semantic representations described above, mirrors the dichotomy in the verb lexicon between "result" and "means/manner"—a basic distinction in the verb lexicon characterized by Levin and Rappaport Hovav [38, 39], and others, and supported by data on language acquisition as described by a number of researchers [5, 6, 23, 25, inter alia]. Following this framework, our semantic representation is characterized by the following basic templates:[24]

(28) (i) **Telic Verbs:**

$$\begin{array}{l}
\text{(a) } \underline{[\text{CAUSE (X,}} \\
\quad [\text{GO}_{\text{Ident}} \text{ (Y,} \\
\qquad [\text{TOWARD}_{\text{Ident}} \text{ (Y, } [\text{AT}_{\text{Ident}} \text{ (Y, } [\langle\text{STATE}\rangle])])])]\underline{)]} \\
\text{(b) } \underline{[\text{CAUSE (X,}} \\
\quad [\text{GO}_{\text{Loc}} \text{ (Y,} \\
\qquad [\langle\text{DIRECTION}\rangle_{\text{Loc}} \text{ (Y, } [\text{AT}_{\text{Loc}} \text{ (Y, Z)}])])]\underline{)]} \\
\text{(c) } \underline{[\text{CAUSE (X,}} \\
\quad [\text{GO}_{\text{Loc}} \text{ (Y,} \\
\qquad [\text{TOWARD}_{\text{Loc}} \text{ (Y, } [\text{AT}_{\text{Loc}} \text{ (Y, } [\langle\text{POSITION}\rangle])])])]\underline{)]}
\end{array}$$

(ii) **Atelic Verbs:**

$$\begin{array}{l}
\text{(a) } [\text{ACT}_{\text{Loc/Perc}} \text{ (X} \\
\qquad [\text{ON}_{\text{Loc/Perc}} \text{ (Y)], [BY } \langle\text{MEANS/MANNER}\rangle])] \\
\text{(b) } [\text{GO}_{\text{Loc/Perc}} \text{ (X,} \\
\qquad [\text{BY } \langle\text{MEANS/MANNER}\rangle])]
\end{array}$$

The underlined portions in (28) are optionally available as a part of the lexical representation. For example, the templates in (28)(i) represent six lexical forms: three with CAUSE and three without.[25]

Figure 1 gives telic and atelic examples of verbs in five broad semantic categories along with their associated LCS representations, based on the telicity analysis from Sections 3 and 4. The distinction between the result and means/manner verbs is reflected in the constant of the LCS representation. For result verbs, the constant corresponds to a resulting position (e.g., ⟨UP⟩ in *lift*), state (e.g., $\langle\text{CLEAN}\rangle_{\text{Ident}}$ in *clean*), or direction (e.g., $\langle\text{AWAY-FROM}\rangle_{\text{Loc}}$ in *escape*). For means/manner verbs, the constant specifies a means (e.g., ⟨FUNNEL⟩ in *funnel*) or a manner (e.g., ⟨SWEEP⟩ in *sweep*). The next section discusses the importance of the telic/atelic distinction for lexical choice in multilingual generation.

[24] We henceforth provide a simplified LCS format, where the type specifications (Event, State, Path, etc.) are omitted (cf. (23) and (24) above). In addition, for the purpose of this discussion, we restrict our attention to non-stative verbs. States, however, are also allowed through the use of the BE primitive.

[25] These templates are a refined version of the ones presented by Dorr and Voss [18]; certain modifications have been made in accordance with our telicity results, e.g., elimination of simultaneously occurring path and means/manner constituents as well as removal of CAUSE from atelic templates.

Category	**Verb**	**Class**	**LCS**	**Template**
Putting	*lift*	9.4 [+telic]	[CAUSE (X, [GO$_{\mathrm{Loc}}$ (Y, [TOWARD$_{\mathrm{Loc}}$ (Y, [AT$_{\mathrm{Loc}}$ (Y, [⟨UP⟩])])])])]	(28)(i)(c)
	funnel	9.7 [∅telic]	[ACT$_{\mathrm{Loc}}$ (X, [ON$_{\mathrm{Loc}}$ (Y)], [BY ⟨FUNNEL⟩])]	(28)(ii)(a)
Removal	*clean*	10.3 [+telic]	[CAUSE (X, [GO$_{\mathrm{Ident}}$ (Y, [TOWARD$_{\mathrm{Ident}}$ (Y, [AT$_{\mathrm{Ident}}$ (Y, [⟨CLEAN⟩$_{\mathrm{Ident}}$ ([⟨OF⟩$_{\mathrm{Poss}}$ (Z)])])])])])]	(28)(i)(a)
	sweep	10.4.1 [∅telic]	[ACT$_{\mathrm{Loc}}$ (X, [ON$_{\mathrm{Loc}}$ (Y)], [BY ⟨SWEEP⟩])]	(28)(ii)(a)
Separation	*separate*	23.1 [+telic]	[CAUSE (X, [GO$_{\mathrm{Ident}}$ (Y, [TOWARD$_{\mathrm{Ident}}$ (Y, [AT$_{\mathrm{Ident}}$ (Y, [⟨SEPARATED⟩$_{\mathrm{Ident}}$ ([⟨FROM⟩$_{\mathrm{Poss}}$ (Z)])])])])])]	(28)(i)(a)
	split	23.2 [∅telic]	[ACT$_{\mathrm{Loc}}$ (X, [ON$_{\mathrm{Loc}}$ (Y)], [BY ⟨SPLIT⟩])]	(28)(ii)(a)
Sound	*say*	37.7 [+telic]	[CAUSE (X, [GO$_{\mathrm{Ident}}$ (Y, [TOWARD$_{\mathrm{Ident}}$ (Y, [AT$_{\mathrm{Ident}}$ (Y, [⟨SAID⟩$_{\mathrm{Ident}}$])])])])]	(28)(i)(a)
	shout	37.3 [∅telic]	[ACT$_{\mathrm{Perc}}$ (X, [ON$_{\mathrm{Perc}}$ (Y)], [BY ⟨SHOUT⟩])]	(28)(ii)(a)
Motion	*escape*	51.1 [+telic]	[GO$_{\mathrm{Loc}}$ (Y, [⟨AWAY-FROM⟩$_{\mathrm{Loc}}$ (Y, [AT$_{\mathrm{Loc}}$ (Y, Z)])])]	(28)(i)(b)
	run	51.3.1 [∅telic]	[GO$_{\mathrm{Loc}}$ (Y, [BY ⟨RUN⟩])]	(28)(ii)(b)

Figure 1. LCSs Based on Telicity Analysis

6. Use of LCS Representations for Multilingual Generation

The representations and constraints described above serve as the foundation for a large-scale, LCS-based lexicon for interlingual MT of English, Arabic, French, Korean, and Spanish. Our interlingual model assumes that the analysis phase—from source-language (SL) input to interlingua (IL)—involves the SL lexicon and the generation phase—from IL form to target-language (TL) output—involves the TL lexicon. The role of the LCSs in each of the lexicons is slightly different, as the task in each direction is inherently distinct: During analysis, the appropriate LCSs are selected from the SL lexicon, as a function of syntax, and the IL is constructed compositionally from the predicate/constant information in the LCSs; During generation, the LCS entries of the TL are accessed and the TL words are chosen such that their LCSs semantically cover the predicate/constant information in the IL. The remainder of this section describes the generation task in more detail, in particular, illustrating the role of the LCS as the primary mediating structure in the two interdependent tasks of *lexical choice* and *syntactic realization*.

6.1. LCS as a Mediator Between Lexical Choice and Syntactic Realization

Our IL formalism, based on the adapted form of the LCS, preserves both semantic structure and semantic content—compositionally derived from the SL sentence—throughout the entire translation process. For example, the sentence *John ran into the room* corresponds to an IL that captures the structure of the motion (i.e., GO/TOWARD/IN) and the content of the manner (i.e., the ⟨RUN⟩):

(29) [CAUSE [+telic]
([$\mathrm{GO}_{\mathrm{Loc}}$ (⟨JOHN⟩, [BY ⟨RUN⟩])],
[$\mathrm{GO}_{\mathrm{Loc}}$ (⟨JOHN⟩,
[⟨TOWARD⟩$_{\mathrm{Loc}}$ (⟨JOHN⟩, [$\mathrm{IN}_{\mathrm{Ident}}$ (⟨JOHN⟩, ⟨ROOM⟩)])])])]

Roughly, the representation above means "John made himself run in such a way that he went into the room." Implicit in the machinery described here is the fundamental assumption—taken from Levin's principles of syntactic realization [38]—that all TL open-class lexical items (e.g., *go*, *run*, etc.) in the surface sentence must be related to a constant in the IL form. The constant comprises the semantic content of the surface sentence, in the spirit of Grimshaw [24] and Levin [38]. Possible target-language lexical items are chosen according to a match with the constant primitive. These are further constrained according to their semantic structure [24, 38], i.e., the non-constant primitives and argument structure.

The key benefit to this structure/constant preservation is that, during generation, the processes of lexical choice and syntactic realization are left open into the TL phase. Thus, TL-specific pragmatic information may be used and stylistic choices may be made in the final generation steps—after the TL lexical options have been identified from the IL form. For the current example, preserving this information allows any number of surface structures to be generated: *John ran into the room*, *John entered the room running*, *John got into the room by running*, etc.

To clarify the nature of the process that composes lexical items into an IL representation, consider the following lexical entries:

(30) (i) **run, running, by_running**: (51.3.1 [∅telic])
[$\mathrm{GO_{Loc}}$ (Y, [BY ⟨RUN⟩])]

(ii) **enter, get_into, into**: (51.1 [+telic])
[$\mathrm{GO_{Loc}}$ (Y, [$\langle\mathrm{TOWARD}\rangle_{\mathrm{Loc}}$ (Y, [$\mathrm{AT_{Loc}}$ (Y, Z)])])]

(iii) **John**: ⟨JOHN⟩ (noun)

(iv) **room**: ⟨ROOM⟩ (noun)

These lexical items serve as the basis of the IL in (29). Their combination adheres to a rule called *Free Composition*, adopted from the LST approach of Levin and Rappaport Havov [38, p. 8], which states: "Templates may be freely augmented up to other possible templates." In particular, since the X position of the Telic Verb templates in (28)(i) can itself be an event (i.e., GO or ACT), a legal augmentation is one in which two subevents are combined, as follows:

(31) [CAUSE [+telic]
([GO/ACT [∅telic]...[BY ⟨MEANS/MANNER⟩]...],
[GO [+telic]...])]

In the sentence *John ran into the room*, the verb *run* corresponds to the [GO/ACT [∅telic]...[BY ⟨MEANS/MANNER⟩]...] subevent and the preposition *into* corresponds to the [GO [+telic]...] subevent.[26] The use of the augmented template is consistent with a line of research most recently articulated by Levin and Rappapaport Havov [38], following other researchers [20, 49, 53, 54, 70], in which causative (telic) verbs (i.e., accomplishments) are comprised of an atelic (activity) causing event and a telic (achievement) event that names a resulting state.

Figure 2 illustrates three possible mappings between surface sentences and the IL form in (29). Note that, in (a)–(c) of Figure 2, the bold-faced constituents are associated with a single constant in the IL representation. The structural positioning of these items in the surface

[26] Illegal combinations such as ∗*John ran entered the room* or ∗*John into the room by running* would be ruled out at the syntactic level, not at the lexical-semantic level; i.e., they would be barred by the structural restriction that each sentential clause contain one, and only one, verb. Grammatical constraints of this type are beyond the scope of this paper.

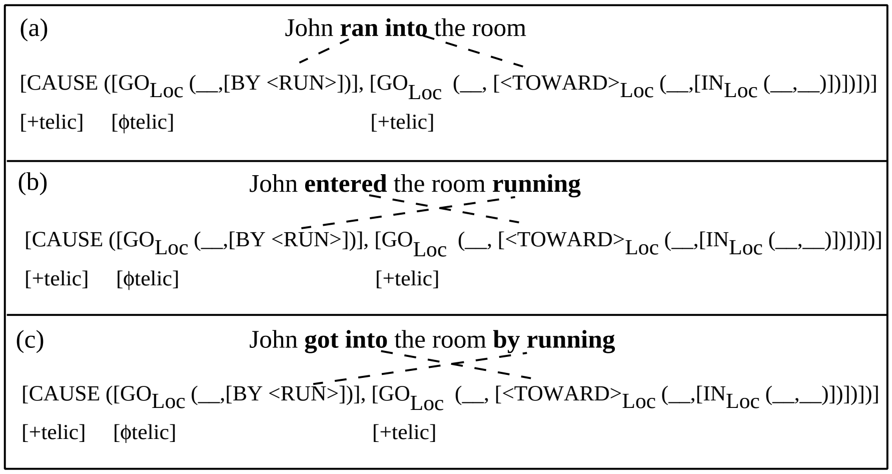

Figure 2. Application of the Free Composition Rule: Relating Lexical Entries to Components of the Surface Sentence

sentence is variable as long as the resulting IL adheres to the Free Composition requirement, i.e., in this example, the structural specification given in (31). This simplified diagram does not illustrate the structural positioning of the arguments *John* and *room*. However, unlike the boldfaced constituents, these are more rigidly positioned by *linking rules* (see, e.g., [32, 39]) that map syntactic arguments into semantic positions, and vice versa. While the statement of such rules is the subject of much debate (and well beyond the scope of this paper), we may assume, for our purposes, that these constituents occupy parallel positions in the surface sentence (i.e., the subject and object positions in the syntactic structure) and the IL (i.e., the logical subject and object, indicated with underscores (__)).

Within this framework, generation from the IL form relies on the same notion of Free Composition used during the analysis phase, where augmented templates must be reduced to the participating lexical items in the IL. Consider generating an English sentence corresponding to the following IL form:

$$
\begin{array}{l}
\text{(32) [CAUSE [+telic]} \\
\quad ([\mathrm{ACT_{Loc}}\ (\langle \mathrm{JOHN} \rangle, [\mathrm{ON_{Loc}}\ (\langle \mathrm{DRIVEWAY} \rangle)], [\mathrm{BY}\ \langle \mathrm{SWEEP} \rangle])], \\
\quad\ [\mathrm{GO_{Ident}} \\
\quad\quad (\langle \mathrm{DRIVEWAY} \rangle, \\
\quad\quad\ [\mathrm{TOWARD_{Ident}} \\
\quad\quad\quad (\langle \mathrm{DRIVEWAY} \rangle, [\mathrm{AT_{Ident}}\ (\langle \mathrm{DRIVEWAY} \rangle, [\langle \mathrm{CLEAN} \rangle_{\mathrm{Ident}}])])])])]
\end{array}
$$

Roughly, this representation means "John acted on the driveway by sweeping such that the driveway became clean." Using this IL form, the generator must first choose an English lexical item (i.e., an LCS template from the English lexicon) and then produce, on the basis of the argument-taking and telicity properties of the chosen item, a legal syntactic realization. Assume, for the sake of argument, that the English sentence that most closely matches this IL representation is *John swept the driveway clean*.

In the case of example (32), the five possible TL lexical items are:

(33) (i) **clean**: (10.3 [+telic])
$[\mathrm{CAUSE}\ (\mathrm{X},$
$[\mathrm{GO}_{\mathrm{Ident}}\ (\mathrm{Y},$
$[\mathrm{TOWARD}_{\mathrm{Ident}}\ (\mathrm{Y},$
$[\mathrm{AT}_{\mathrm{Ident}}\ (\mathrm{Y}, [\langle\mathrm{CLEAN}\rangle_{\mathrm{Ident}}\ ([\langle\mathrm{OF}\rangle_{\mathrm{Poss}}\ (\mathrm{Z})]))])])])]$[27]

(ii) **sweep, by_sweeping**: (10.4.1 [$\emptyset$telic])
$[\mathrm{ACT}_{\mathrm{Loc}}\ (\mathrm{X}, [\mathrm{ON}_{\mathrm{Loc}}\ (\mathrm{Y})], [\mathrm{BY}\ \langle\mathrm{SWEEP}\rangle])]$

(iii) **swept**: (10.3 [+telic])[28]
$[\mathrm{CAUSE}\ (\mathrm{X},$
$[\mathrm{GO}_{\mathrm{Ident}}\ (\mathrm{Y},$
$[\mathrm{TOWARD}_{\mathrm{Ident}}\ (\mathrm{Y},$
$[\mathrm{AT}_{\mathrm{Ident}}\ (\mathrm{Y}, [\langle\mathrm{SWEEP}\rangle_{\mathrm{Ident}}\ ([\langle\mathrm{OF}\rangle_{\mathrm{Poss}}\ (\mathrm{Z})]))])])])]$

(iv) **John**: ⟨JOHN⟩ (noun)

(v) **driveway**: ⟨DRIVEWAY⟩ (noun)

There are multiple combinations of these five words that might make sense, for example:

(34) (i) John cleaned the driveway by sweeping

(ii) ∗John cleaned the driveway swept

(iii) John swept the driveway clean

The question to ask is how (34)(ii) is ruled out as a TL sentence. At first glance, it appears that the [+telic] marker in (32) would force the lexical

[27] The $[\langle\mathrm{OF}\rangle_{\mathrm{Poss}}$ ...] substructure in the definition of *clean* has an extra set of parentheses; this component is optionally realized in the surface structure, i.e., *John cleaned the driveway of leaves* vs. *I cleaned the driveway*.

[28] Although *sweep* and *clean* are both REMOVE VERBS, *sweep* differs from *clean* in that it does not fall into the same semantic class as its adjectival counterpart. That is, both the verbal and adjectival senses of *clean* are CLEAR VERBS (class 10.3) whereas the verb *sweep* and the adjective *swept* are in two different classes, WIPE VERBS (class 10.4.1) and CLEAR VERBS (class 10.3), respectively. This is because the verb *sweep* is an activity, whereas the adjectival form *swept*, like *clean*, describes a resulting state.

entry for *clean*, (33)(i), to be selected as the main verb of the output sentence. However, if this verb is chosen, the only way to incorporate the ⟨SWEEP⟩ component of meaning into the output sentence is as an instrumental phrase (i.e., *by sweeping*) which occurs in the activity template (33)(ii). If this form is selected, the combination of the two lexical entries fully covers the structure of the IL and sentence (34)(i) is produced. Sentence (34)(ii), on the other hand, would be ruled out because the resultative form of *sweep* in (33)(iii) is in a structure that does not occur as a substructure of the IL. As for sentence (34)(iii), the lexical entry for *sweep*, (33)(iii), would be selected as the main verb, and in this case, the remainder of the IL exactly matches the lexical entry for *clean*, (33)(i). Thus, the sentence (34)(iii) is a possible TL sentence. Note that the adjectival form *clean* in the surface sentence derives from the constant ⟨CLEAN⟩ in the template for the verb *clean*.

The only question remaining is whether the selections made for (34)(i) and (iii) are appropriate with respect to telicity. For example, the CAUSE event in (32) is [+telic] and the *sweep/by_sweeping* template is [∅telic]. Recall from section 2 that the privative model of aspect [45, 46] allows verbs that are unmarked (i.e., [∅telic]) in a lexical entry to become marked (i.e., [+telic]) when the verb is combined with other constituents. The presence of the [+telic] marker in the lexical entry for *clean* (in (33)(i)) combines with the [∅telic] marker in the lexical entry for *sweep* (in (33)(ii)) such that a [+telic] marker occurs on the overall IL representation. Thus, both sentences (34)(i) and (iii) are available as possible surface sentences for the IL in (32).

We now consider this framework for generation of TL sentences in another language, in particular, Spanish, in which the change-of-state resultative construction (e.g., *John swept the driveway clean*) is not available. Rather, the appropriate TL sentence would be either a paraphrase with two clauses or a single clause with some omitted or redistributed information:

(35) (i) John barrió el camino de entrada y lo dejó limpio.
'John swept the driveway and left it clean.'

(ii) John limpió el camino de entrada (con una escoba).
'John cleaned the driveway (with a broom).'

For each natural language, we identify allowable sets of syntactic-realization operations in that language, and then we identify, more narrowly, the classes of lexical items to which these operations apply. In this case, we parameterize the IL/syntax mapping so that the change-of-state resultative operation is applied in English, but not in Spanish. The parameterization permits each language to have its own idiosyncratic syntactic realization while sharing underlying IL representations.

The first sentence above most closely reflects the meaning of the English change-of-state resultative construction since the verb *barrer* incorporates the *sweeping* component of meaning and the clause containing *dejar limpio* ('to leave clean') reflects the resulting state. However, the second translation, the more "natural" of the two according to our native informants, is not as straightforward to generate as the first one—with or without the instrumental phrase *con una escoba* ('with a broom'). The preference to translate the English atelic verb *sweep* as the Spanish telic equivalent *limpiar* is consistent with our earlier prediction that such languages may be said to have (or perhaps employ) more [+telic] verbs.

We now examine these two cases in more detail. The relevant TL lexical items for the sentences above are:

(36) (i) **barrer**: (10.4.1 [∅telic])
$[\mathrm{ACT_{Loc}}\ (\mathrm{X}, [\mathrm{ON_{Loc}}\ (\mathrm{Y})], [\mathrm{BY}\ \langle\mathrm{SWEEP}\rangle])]$

(ii) **limpiar, dejar_limpio**: (10.3 [+telic])
$[\mathrm{CAUSE}\ (\mathrm{X},$
$[\mathrm{GO_{Ident}}\ (\mathrm{Y},$
$[\mathrm{TOWARD_{Ident}}\ (\mathrm{Y}, [\mathrm{AT_{Ident}}\ (\mathrm{Y}, \langle\mathrm{CLEAN}\rangle_{\mathrm{Ident}})])])])]$[29]

(iii) **con**: $[\langle\mathrm{WITH_{Instr}}\ (\mathrm{Z})]$

(iv) **John**: ⟨JOHN⟩ (noun)

(v) **escoba**: ⟨BROOM⟩ (noun)

(vi) **camino_de_entrada**: ⟨DRIVEWAY⟩ (noun)

In order to produce a sentence with the same telicity value as the IL in (32), two lexical items must be combined (*barrer* and *dejar_limpio*), as in the English case above, while taking into account parameter settings necessary for producing grammatical Spanish syntactic structures. Since the resultative construction is barred in Spanish, the two-clause sentence (35)(i) is produced from the *barrer* and *dejar_limpio* templates, where the former fills the Y position in the latter.[30] As in the English case, the adjectival form *limpio* in the surface sentence derives from the constant ⟨CLEAN⟩ in the template for the verb construction *dejar limpio*. The difference is that the parameterization restriction

[29] Note that, unlike the English lexical entries for *clean* and *swept* the $[\langle\mathrm{OF}\rangle_{\mathrm{Poss}}\ \ldots]$ substructure does not occur in the Spanish lexical entries. This is difference is due to the fact that adjectives like *clean* cannot be used in an analogous *of* construction in Spanish: *John barrió el camino de entrada ∗de hojas* ('John swept the driveway of leaves').

[30] As might be expected, the sentence *John barrió el camino de entrada y lo limpió* is also produced as a TL sentence. According to our informants, this sentence is viewed on a par with (35)(i) in terms of acceptability, i.e., it is acceptable, but not as natural sounding as (35)(ii).

against the resultative forces the template containing the ⟨CLEAN⟩ constant to be realized in a secondary clause (conjoined by *y*).[31] Thus, sentence (35)(i) is easily produced from the lexical entries above.

Case (35)(ii), on the other hand, could only be generated from one of the following IL forms:

(37) (i) [CAUSE [+telic]
([$\mathrm{ACT_{Loc}}$ (⟨JOHN⟩,
[$\mathrm{ON_{Loc}}$ (⟨DRIVEWAY⟩)], [⟨$\mathrm{WITH_{Instr}}$ (⟨BROOM⟩)])],
[$\mathrm{GO_{Ident}}$
(⟨DRIVEWAY⟩,
[$\mathrm{TOWARD_{Ident}}$
(⟨DRIVEWAY⟩, [$\mathrm{AT_{Ident}}$ (⟨DRIVEWAY⟩, [⟨CLEAN⟩$_{\mathrm{Ident}}$])])])])]

(ii) [CAUSE [+telic]
(⟨JOHN⟩,
[$\mathrm{GO_{Ident}}$
(⟨DRIVEWAY⟩,
[$\mathrm{TOWARD_{Ident}}$
(⟨DRIVEWAY⟩, [$\mathrm{AT_{Ident}}$ (⟨DRIVEWAY⟩, [⟨CLEAN⟩$_{\mathrm{Ident}}$])])])])]

Note that, in both of these cases, the second subevent exactly matches its counterpart in (32). However, in (37)(i), the [⟨$\mathrm{WITH_{Instr}}$ (⟨BROOM⟩)] subcomponent conflicts with [BY ⟨SWEEP⟩] in the IL and, in (37)(ii), ⟨JOHN⟩ conflicts with the entire the entire first subevent in the IL. Thus, we are left with the question of whether these alternate structures—which actually match a major subcomponent of the IL representation—are adequate matches for translation into Spanish.

Currently, our system opts for the more literal translation over a loss, or modification, of information in the final output. However, it is clear that such cases must be addressed since certain languages may not have an equivalent to either the resultative construction (as in (34)(iii)) or the result-type verb (as in (35)(i)).[32] This type of mismatch is addressed further in the next section.

6.2. Additional Guidance for Lexical Choice: KR

The entire representational framework outlined above is designed to integrate with a KR-based processing module described in [18]. Since this module was originally developed with only telic verbs in mind,

[31] The ordering of these two conjoined clauses in the TL sentence reflects the expected temporal ordering, i.e., the second subevent is viewed as a result of completing the first subevent.

[32] We are indebted to a reviewer who pointed this out for a Russian translation of *shovel clean*.

extensions are currently underway to incorporate atelic verbs and to provide constraints on their combination with telic verbs. In our approach, using the LST-based structures for the IL level of representation, the full IL form is the guiding structure out of which lexical items are chosen and syntactically combined with other lexical items. The KR-based module provides additional guidance by determining the range of TL lexical items that can "cover" the full IL form using a constrained search mechanism. Each SL and TL lexicon is tied into a larger "lexicalization space" called the *IL Lexicon*.

Following the design in [18], the IL Lexicon provides two types of links, shown in Figure 3: (1) a *reduction* link (dotted arrow), which relates a causative form to its non-causative counterpart; (2) a *subsumption* link (solid arrow), which relates a lexical IL form to its more specific counterpart. The lexical entries (Spanish *forzar, mover, levantar, mudar, levantarse* and English *cause*, *move*, and *lift*) are organized hierarchically by automatic means. (See [19] for details.) Note that, unlike the English words (such as *move*, which occurs as both a causative and non-causative lexical item), the Spanish words are not "reused" here. In general, verbs are not as heavily overloaded in Spanish as they are in English; Spanish non-causatives are distinguished from their causative counterparts by means of distinct lexical items or by the use of a reflexive "se".

The IL Lexicon is designed to handle lexical mismatches during the lexical choice process. In particular, the generation algorithm is designed to navigate through the IL Lexicon, following reduction links (dotted arrows) and subsumption links (solid arrows), to compensate for cases where the appropriate lexical items in one language do not line up precisely with the lexical items in another language. In the case given in Figure 3, there are no mismatches, i.e., every verb in English is paired with a Spanish verb, and vice versa. We will look at a more complicated case shortly, i.e., one where lexical items are not precisely lined up between English and Spanish.

The process of lexical choice uses an augmented version of the lexicalization space in [18], where reduction links cross the line between atelic and telic verbs as illustrated in Figure 4. The main augmentation to the original IL Lexicon framework is that reduction links are now allowed to refer to either of the two argument positions in a causative (telic) template. For example, the reduction link between the top-most CAUSE template and the atelic template for *sweep* and *barrer* is one that resolves the *first* argument of the causative form; by contrast, the reduction link between the template for English *make* and English *become* is one that resolves the *second* argument of the causative form. This extension is entirely consistent with the Free Composition require-

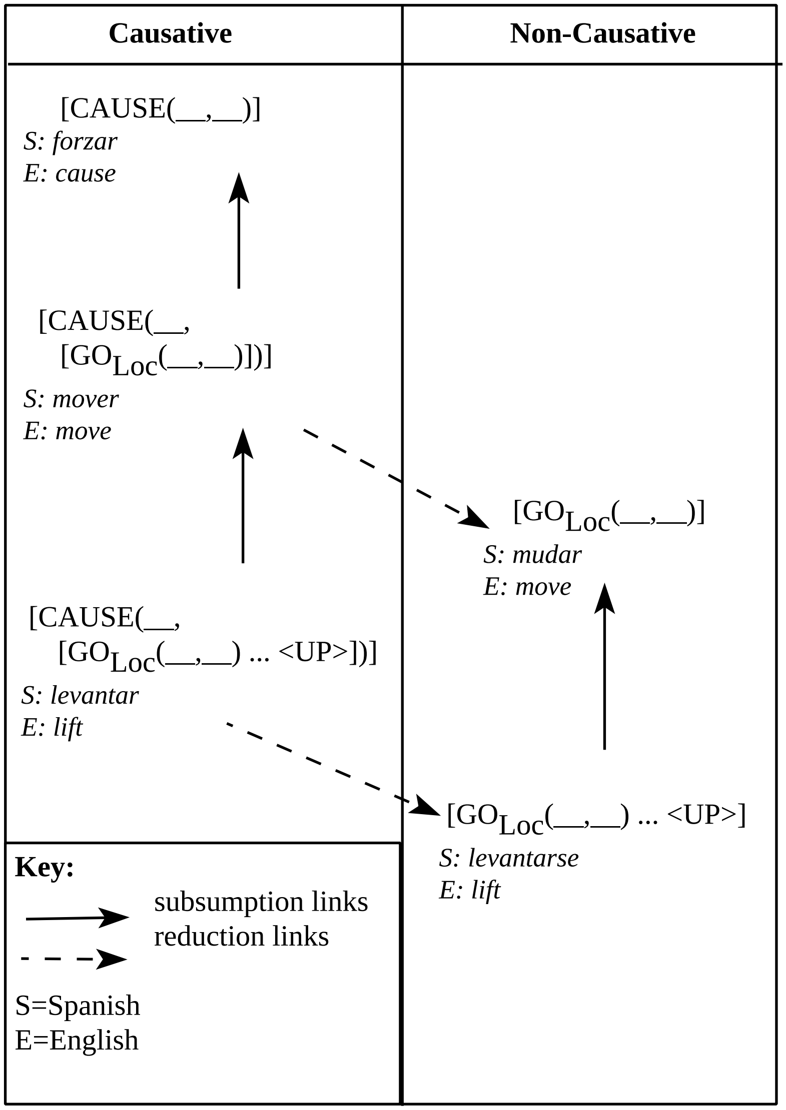

Figure 3. LCSs with Directed Subsumption Links and Crossover Reduction Links in IL Lexicon: Spanish and English

ment stated in the previous section, i.e., the reduction links conform to the structural restrictions identified in (31), where the causative (telic) template contains an atelic causing event and a telic event that names a resulting state. The benefit of using the reduction links for resolving both argument positions is that the same navigation algorithm (outlined in [18]) may be used for selection and realization of appropriate TL lexical items *across* telicity categories.

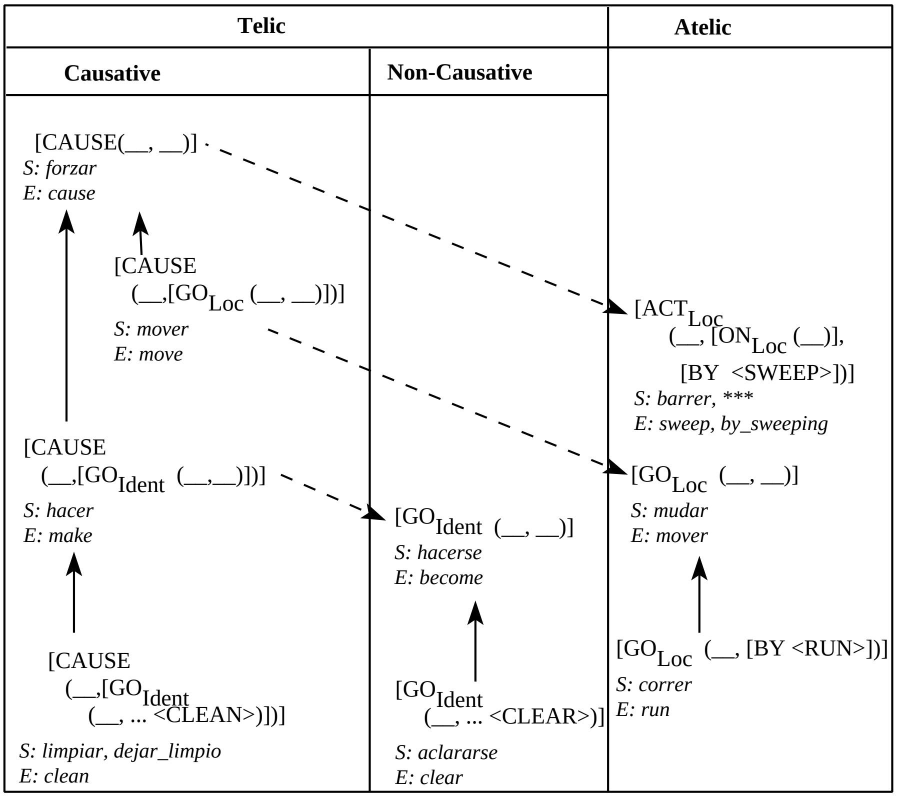

Figure 4. Extension to IL Lexicon Framework: Using Reduction Links to Distinguish Between Telic and Atelic Verbs

Returning to our *sweep/barrer* example from the previous section, we now reconsider the case of generating a surface-language sentence for the IL form in (32), repeated here:

(38)
$$
\begin{array}{l}
[\mathrm{CAUSE}\ [+\mathrm{telic}] \\
\quad ([\mathrm{ACT_{Loc}}\ (\langle \mathrm{JOHN} \rangle, [\mathrm{ON_{Loc}}\ (\langle \mathrm{DRIVEWAY} \rangle)], [\mathrm{BY}\ \langle \mathrm{SWEEP} \rangle])], \\
\quad\ [\mathrm{GO_{Ident}} \\
\quad\quad (\langle \mathrm{DRIVEWAY} \rangle, \\
\quad\quad\ [\mathrm{TOWARD_{Ident}} \\
\quad\quad\quad (\langle \mathrm{DRIVEWAY} \rangle, [\mathrm{AT_{Ident}}\ (\langle \mathrm{DRIVEWAY} \rangle, [\langle \mathrm{CLEAN} \rangle_{\mathrm{Ident}}])])])])]
\end{array}
$$

Unlike the ontology in Figure 3, the lexical items in Figure 4 do not line up precisely, as symbolized by the '***' notation marking the lexical gap. In the current example, this notation indicates that the English phrase *by sweeping* has no equivalent in Spanish.

The translation between the source and target language for this example involves a mapping between a telic construction in English (i.e., the atelic verb *sweep* along with a telicizing adjective *clean*) and a telic construction in Spanish (i.e., the atelic verb *barrer* along with a telicizing construction *dejar limpio*, or else just the single telic verb *limpiar*). The following steps are taken:

- The IL in (38) matches the Spanish verb entry for *dejar_limpio* (and also *limpiar*) with the exception of an unspecified causing subevent (i.e., an empty first argument).

- The position corresponding to the unspecified causing subevent is resolved by following subsumption links "upwards" and then reduction links "sideways" until a causing subevent is found that matches the atelic component of meaning in the IL.

- The template for *barrer* is found via the reduction link from the top-most CAUSE.

- The *dejar_limpio* and *barrer* templates are combined to produce a structure that successfully matches the IL in (38), and the sentence *John barrió el camino de entrada y lo limpió* is produced.

The cross-language mismatch above is resolved solely by means of navigation through the IL Lexicon. That is, the structure and content of the lexical entries alone are sufficient enough to provide a full cover of the IL. However, there are numerous cases where the IL Lexicon cannot resolve mismatches without additional knowledge about the concepts underlying the constants in the lexical entries. This is precisely why generation of *limpiar*—with or without the instrumental phrase *con una escoba*—from the IL is not as straightforward as the *sweep clean* and *barrer* cases. The sentence *John limpió el camino de entrada con una escoba* must, nevertheless, be considered as a TL possibility, not just because it is more natural sounding than the others, but because it is the *only* possibility in certain languages, e.g., Russian.

The question of whether it is possible to omit or redistribute different pieces of information for this case relies heavily on whether enough world knowledge and context pertaining to the sentence are available to make this decision. For example, if there exists a KR relationship of "canonical instrument" between the ⟨SWEEP⟩ and ⟨BROOM⟩ concepts, as shown in Figure 5, then this relationship can be exploited to produce the TL sentence with the instrumental phase.

Clearly this type of information has no place in the predicates and constants of the IL Lexicon, but our lexical choice algorithm would

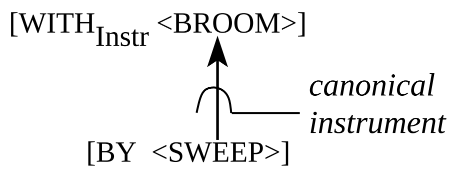

Figure 5. Canonical Instrument Relationship between ⟨SWEEP⟩ and ⟨BROOM⟩

be enhanced significantly if such KR relationships were made available during the navigation process. In this case, the following steps are taken:

- The verb *limpiar* is selected as the main verb since the IL in (38) matches this entry with the exception of an unspecified causing (atelic) subevent.

- As before, the position corresponding to the unspecified causing subevent is resolved by following subsumption links "upwards" and then reduction links "sideways" to select a matching atelic event.

- Unlike the previous example, the inability to produce the *by sweeping* phrase in Spanish (as indicated by the '***' notation) forces the KR relationship to be accessed.

- The [BY ⟨SWEEP⟩] sub-component is replaced by [⟨WITH$_{\text{Instr}}$ (⟨BROOM⟩)].

- The atelic component (with the replaced sub-component) is combined with the telic entry for *limpiar*.

- The [⟨WITH$_{\text{Instr}}$ (Z)] corresponds to the entry for *con* and ⟨BROOM⟩ corresponds to the entry for *escoba*; thus, the Spanish sentence *John limpió el camino de entrada con una escoba* is produced.

Alternatively, if one can determine from the context of the sentence that the activity causing the clean state of the driveway involves a broom, then perhaps the *sweeping* component of the activity may be dropped in the TL output. In this case, the focus of the sentence would be on the resulting state (i.e., *clean*). The omission of such components of meaning has been discussed at length by Slobin [58], where English and Spanish are distinguished by the degree to which each language includes context information. Spanish typically includes more contextual background than English and, as such, does not employ verbs that convey manner of motion as often as English does. In fact, according to

Slobin, Spanish translators often make changes that *reduce* the amount of information conveyed in the English sentence in order to produce the Spanish sentence. It is clear that, if our lexical choice machinery were augmented with a richer theory of context—coupled with constraints on which components of meaning may be legally omitted in the TL sentence—these reduced cases could be adequately handled.

7. Conclusion

We have investigated the extent to which lexicalization patterns involving the lexical aspect feature [+telic] may be exploited in the representation of Lexical Conceptual Structure to facilitate the task of lexical choice and syntactic realization. The main contribution of this work is a means for generating a TL sentence that preserves both the semantic structure and the semantic content of the IL, by defining the LCS using the LST framework. Approaches that ignore the structural nature of lexical items and their combination stand to lose the benefit of regularities that exist *within* and *across* languages in the lexicon and in the surface syntactic form. By decomposing verbs into semantic *structure* and semantic *content*, we exploit such regularities.

We used the syntactic alternations and semantic verb classes in [37] as a preliminary study of the relationship between lexical aspect features and the semantic features manipulated by the syntactic alternations. Having enriched the data in [37] with telicity markings for the verb classes and alternations, we built a database of English LCS representations for the 3828 verb entries in Levin's book [37], where an *entry* is a "semantic class/verb sense" pair, e.g., "9.4/lift" from Figure 1. These LCSs were automatically derived from a manually-encoded set of templates (taken from (28)) for Levin's semantic classes, consisting of 2775 unique verbs distributed across 3828 verbs in Levin (some verbs occur in multiple classes). We have augmented this database with LCS representations for 3500 verbs not occurring in Levin by associating with each verb a semantic class and then instantiating the relevant LCS template.[33] The full set of LCSs is being used in a large-scale analysis component for a MT system [15, 17, 18]. A prototype version of the generation component has been constructed for complete end-to-end translation [19].

During analysis, the appropriate LCSs in the SL are selected as a function of syntax, and the IL constructed compositionally from the semantic structure and constant information therein. During gen-

[33] For details of experiments in the automatic construction of an LCS-based lexicon for other languages, e.g., Arabic, see [12].

eration, the LCS entries of the TL are accessed, and lexical choice is constrained by requiring TL words to cover the semantic structure/constant in the IL in permitted syntactic structures. Areas of future investigation include development of a framework for handling multiple sentences; in particular, a rich theory of context coupled with our rules of composition might allow us to identify cases where the IL underlying the SL sentence must be either reduced or modified in order to produce an appropriate TL sentence.

Acknowledgements

The research reported herein was supported, in part, by Army Research Office contract DAAL03-91-C-0034 through Battelle Corporation, NSF NYI IRI-9357731, Alfred P. Sloan Research Fellow Award BR3336, and a General Research Board Semester Award. We are indebted to three anonymous reviewers and to Beth Levin whose comments provided excellent guidance in producing this final version. Additional commentary on issues pertaining to this research was provided by: Doug Jones, Grace Song, and Clare Voss. Our Spanish informants provided us with numerous examples to aid us in our understanding of telicity-related distinctions between English and Spanish: Irma Amenero, Suni Cabrera, Pablo Gomez, and Stephanie Pringhipakis.

Appendix

A. Analysis of Alternations with Respect to Telicity

This appendix contains our five-way classification of alternations with respect to telicity: (i) alternations that indicate telicity (all participating verbs are [+telic] in their basic sense), (ii) alternations and constructions that add telicity (all participating verbs are [+telic] in the relevant construction), (iii) alternations that indicate atelicity (all participating verbs are [$\emptyset$telic] in their basic sense), (iv) alternations and constructions that are irrelevant with respect to (a)telicity (some participating verbs are [+telic] and others [$\emptyset$telic], and their categorization is not systematically affected by the relevant construction), and, for completeness, (v) a small number of alternations that cannot be classified.

(i) **Alternations Indicating [+telic]:** 2.3.5, 2.4.3, 2.4.4, 2.6, 3.4, 2.5.4.

(ii) **Alternations Adding [+telic]:**[34] 1.1.1, 1.1.2.2, 1.2.1, 1.2.7, 1.4.1, 2.1, 2.2, 2.3.1, 2.3.2, 2.3.3, 2.4.1, 2.4.2, 2.5.1, 2.5.3, 2.5.6, 2.7, 2.9, 2.14, 3.5, 4.2, 7.1, 7.2, 7.4, 7.5, 7.6.1, 7.6.2, 7.8.

(iii) **Alternations Indicating [∅telic]:**[35] 1.1.3, 1.3, 2.8, 2.10, 2.11.3, 2.11.4, 2.11.5, 2.13.3, 3.7, 3.8, 3.10.

(iv) **Alternations for which (A)telicity is Irrelevant:** 1.1.2.1, 1.1.2.3, 1.2.2, 1.2.3, 1.2.4, 1.2.5, 1.2.6.1, 1.2.6.2, 1.2.8, 1.4.2, 2.3.4, 2.5.2, 2.5.5, 2.11.1, 2.11.2, 2.12, 2.13.1, 2.13.2, 2.13.4, 2.13.5, 3.1, 3.6, 3.9, 4.1, 5.1, 5.2, 5.3, 5.4, 6.1, 6.2, 7.3, 7.7, 8.1, 8.2, 8.3, 8.4, 8.5, 8.6.

(v) **Unclassifiable Alternations:** 3.2, 3.3

B. Telic and Atelic Semantic Classes

This appendix lists the semantic classes that fall into the telic and atelic categories, using the classified alternations as positive telicity diagnostics.

(i) **Telic Semantic Classes:** 10.3, 13.4.1, 22.1, 23.1, 23.4, 26.6, 29.1, 29.2, 29.4, 29.5, 36.1, 36.2, 36.3, 37.6, 37.7.

(ii) **Atelic Semantic Classes:** 9.7, 9.10, 12, 10.1, 10.2, 10.4.1, 10.4.2, 10.5, 13.7, 14, 17.1, 17.2, 18.1, 18.2, 18.3, 19, 20, 21.1, 21.2, 22.3, 22.4, 24, 25.1, 25.2, 25.4, 26.1, 26.2, 26.3, 26.5, 26.7, 29.8, 31.1, 31.2, 35.1, 35.2, 35.3, 35.5, 35.6, 37.3, 38, 39.1, 39.2, 39.3, 39.4, 39.5, 39.6, 40.1.1, 40.2, 40.3.2, 40.7, 40.8.2, 40.8.3, 41.2.1, 43.2, 43.4, 45.1, 45.2, 45.3, 45.4, 47.2, 47.7, 47.8, 49, 51.3.1, 51.3.2, 51.4.1, 51.4.2, 51.5, 53.2, 54.1, 23.2.

C. Additional Telic and Atelic Semantic Classes

This appendix lists the semantic classes that fall into the telic and atelic categories, using the classified alternations as negative telicity diagnostics. Manual inspection of these classes confirms the assignment of 31 of 45 classes on the relevant interpretations, with exceptions annotated

[34] We assume that alternations 7.6.1 and 7.6.2 serve as atelic diagnostics only if the verb *cut* is not not classified as [+telic]

[35] We have divided the *Search* (2.11) into five variants. Suffixes 1–5 (e.g., 2.11.3) have therefore been added to make this distinction.

with the - symbol. Those marked 'm' are mixed in telicity, with classification based on an alternation that applies to a uniformly (a)telic subset of the relevant class.

(i) **Additional Telic Semantic Classes:** 9.1, 9.2, 9.3, 9.4, 9.5, 9.6 (m), 9.8, 9.9, 10.6, 10.7, 10.8, 11.1, 11.2 (m), 11.3, 11.4 (-), 13.1, 13.2, 13.5.1, 15.2 (-), 18.4, 26.4, 30.1 (-), 30.2, 31.4 (-), 32.1 (-), 33 (-), 34 (-), 35.4, 40.1.2 (-), 40.1.3 (m), 40.8.1 (-), 41.1.1, 41.2.2, 42.1, 42.2, 43.1 (-), 43.3 (-), 44, 47.3 (-), 48.1.1, 48.1.2, 50, 51.1 (-).

(ii) **Additional Atelic Semantic Classes:** 22.5, 23.3 (-).

References

1. S. Ali and S. Shapiro. Natural Language Processing Using a Propositional Semantic Network with Structured Variables. *International Journal of Minds and Machines, Special Issue on Knowledge Representation for Natural Language*, 3:421–451, 1993.
2. J. Anderson. *The Grammar of Case: Towards a Localist Theory.* Cambridge University Press, Cambridge, England, 1971.
3. J.D. Apresjan. Regular Polysemy. *Linguistics*, 142:5–32, 1973.
4. Edmond Bach. The Algebra of Events. *Linguistics and Philosophy*, 9:5–16, 1986.
5. Douglas A. Behrend. The Development of Verb Concepts: Children's Use of Verbs to Label Familiar and Novel Events. *Child Development*, 61:681–696, 1990.
6. Douglas A. Behrend, L. Lynn Harris, and Kelly B. Cartwright. Morphological Cues to Verb Meaning: Verb Inflections and the Initial Mapping of Verb Meanings. *Journal of Child Language*, 22:89–106, 1995.
7. Laurel J. Brinton. *The Development of English Aspectual Systems: Aspectualizers and Post-Verbal Particles.* Cambridge University Press, Cambridge, 1988.
8. D. Alan Cruse. Polysemy and Related Phenomena from a Cognitive Linguistic Viewpoint. In P. Saint-Dizier and E. Viegas, editors, *Computational Lexical Semantics*, pages 33–49. Cambridge University Press, England, 1995.
9. Chrysanne DiMarco, Graeme Hirst, and Manfred 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.
10. Bonnie J. Dorr. A Parameterized Approach to Integrating Aspect with Lexical-Semantics for Machine Translation. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, pages 257–264, University of Delaware, Newark, DE, 1992.
11. Bonnie J. Dorr. *Machine Translation: A View from the Lexicon.* The MIT Press, Cambridge, MA, 1993.
12. Bonnie J. Dorr, Joseph Garman, and Amy Weinberg. From Syntactic Encodings to Thematic Roles: Building Lexical Entries for Interlingual MT. *Machine Translation*, 9(3), 1995.
13. Bonnie J. Dorr and Doug Jones. Role of Word Sense Disambiguation in Lexical Acquisition: Predicting Semantics from Syntactic Cues. In *Proceedings*

of the International Conference on Computational Linguistics, Copenhagen, Denmark, 1996.

14. Bonnie J. Dorr and Doug Jones. Use of Syntactic and Semantic Filters for Lexical Acquisition: Using WordNet to Increase Precision. In *Proceedings of the Workshop on Predicative Forms in Natural Language and Lexical Knowledge Bases*, Toulouse France, 1996.
15. Bonnie J. Dorr, Dekang Lin, Jye hoon Lee, and Sungki Suh. Efficient Parsing for Korean and English: A Parameterized Message Passing Approach. *Computational Linguistics*, 21(2):255–263, 1995.
16. Bonnie J. Dorr and Martha Palmer. Building a LCS-Based Lexicon in TAGs. In *Proceedings of the AAAI-95 Spring Symposium Series, Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity, and Generativity*, Stanford, CA, March 27–29 1995.
17. Bonnie J. Dorr and Clare Voss. Machine Translation of Spatial Expressions: Defining the Relation between an Interlingua and a Knowledge Representation System. In *Proceedings of Twelfth Conference of the American Association for Artificial Intelligence*, pages 374–379, Washington, DC, 1993.
18. Bonnie J. Dorr and Clare Voss. A Multi-Level Approach to Interlingual MT: Defining the Interface between Representational Languages. *International Journal of Expert Systems*, 1996.
19. Bonnie J. Dorr, Clare Voss, Eric Peterson, and Michael Kiker. Concept Based Lexical Selection. In *AAAI 1994 Fall Symposium on Knowledge Representation for Natural Language Processing in Implemented Systems*, New Orleans, LA, 1994.
20. David Dowty. *Word Meaning in Montague Grammar*. Reidel, Dordrecht, 1979.
21. David Dowty. Toward a Semantic Analysis of Verb Aspect and the English "Imperfective" Progressive. *Linguistics and Philosophy*, 1:45–77, 1991.
22. Howard B. Garey. Verbal Asepct in French. *Language*, 33(2):91–110, 1957.
23. Dedre Gentner. On Relational Meaning: The Acquisition of Verb Meaning. *Child Development*, 49:988–998, 1978.
24. Jane Grimshaw. Semantic Structure and Semantic Content in Lexical Representation. unpublished ms., Rutgers University, New Brunswick, NJ, 1993.
25. Jess Gropen, Steven Pinker, Michelle Hollander, and Richard Goldberg. Syntax and Semantics in the Acquisition of Locative Verbs. *Journal of Child Language*, 18:115–151, 1991.
26. Jeffrey S. Gruber. *Look* and *See*. *Language*, 43(4):937–947, 1967.
27. Mohamed Guerssel, Kenneth Hale, Mary Laughren, Beth Levin, and Josie White Eagle. A Cross-linguistic Study of Transitivity Alternations. In W. H. Eilfort, P. D. Kroeber, and K. L. Peterson, editors, *CLS 21, Part 2*, pages 48–63, 1985.
28. B. Heine, U. Claudi, and F. Huennemeyer. *Grammaticalization: A Conceptual Framework*. University of Chicago Press, Chicago, IL, 1991.
29. Graeme Hirst. Resolving Lexical Ambiguity Computationally with Spreading Activation and Polaroid Words. In S.I. Small, G.W. Cottrell, and M.K. Tanenhaus, editors, *Lexical Ambiguity Resolution*. Morgan Kaufmann, San Mateo, CA, 1988.
30. Lucja Iwanska. Logical Reasoning in Natural Language: It is all about Knowledge. *International Journal of Minds and Machines, Special Issue on Knowledge Representation for Natural Language*, 3:475–510, 1993.
31. Ray Jackendoff. *Semantics and Cognition*. The MIT Press, Cambridge, MA, 1983.
32. Ray Jackendoff. *Semantic Structures*. The MIT Press, Cambridge, MA, 1990.
33. Suzanne Kemmer, editor. *The Middle Voice. Typological Studies in Language 23*. John Benjamins, Amsterdam, 1993.

34. Anthony Kenny. *Action, Emotion and Will.* Routledge and Kegan-Paul, London, 1963.
35. Adam Kilgarriff. Inheriting Polysemy. In P. Saint-Dizier and E. Viegas, editors, *Computational Lexical Semantics*, pages 319–335. Cambridge University Press, England, 1995.
36. R. Langacker. *Foundations of Cognitive Grammar, Vol. 1: Theoretical Prerequisites.* Stanford University Press, Stanford, CA, 1987.
37. Beth Levin. *English Verb Classes and Alternations: A Preliminary Investigation.* University of Chicago Press, 1993.
38. Beth Levin and Malka Rappaport Hovav. The Elasticity of Verb Meaning. In *Proceedings of the Tenth Annual Conference of the Israel Association for Theoretical Linguistics and the Workshop on the Syntax-Semantics Interface*, Bar Ilan University, Israel, 1995.
39. Beth Levin and Malka Rappaport Hovav. *Unaccusativity: At the Syntax-Lexical Semantics Interface.* The MIT Press, Cambridge, MA, 1995. LI Monograph 26.
40. Lori Levin and Sergei Nirenburg. Principles and Idiosyncracies in MT Lexicons. In *Working Notes for the AAAI Spring Symposium on Building Lexicons for Machine Translation*, pages 122–131, Stanford University, CA, 1993.
41. Talke Macfarland. *Cognate Objects and the Argument/Adjunct Distinction in English.* PhD thesis, Northwestern University, Evanston, IL, 1995.
42. A. Mourelatos. Events, Processes, and States. In P. Tedeschi and A. Zaenen, editors, *Syntax and Semantics 14: Tense and Aspect*, pages 191–212. Academic Press, New York, NY, 1981. Originally published in *Linguistics and Philosophy* 2, 1978:415-434.
43. Sergei Nirenburg, Jaime Carbonell, Masaru Tomita, and Kenneth Goodman. *Machine Translation: A Knowledge-Based Approach.* Morgan Kaufmann, San Mateo, CA, 1992.
44. Mari Broman Olsen. Lexical Semantic Typology of Motion Verbs: Insight from Translation. In M. Alexander and M. Dressler, editors, *Papers from the Second Annual Meeting of the Formal Linguistic Society of Mid-America*, pages 255–271. University of Michigan, Ann Arbor, MI, 1991.
45. Mari Broman Olsen. *The Semantics and Pragmatics of Lexical and Grammatical Aspect.* PhD thesis, Northwestern University, Evanston, IL, 1994.
46. Mari Broman Olsen. The Semantics and Pragmatics of Lexical Aspect Features. In *Proceedings of the Formal Linguistic Society of Midamerica V*, University of Illinois, Urbana-Champaign, May 20–22 1994. In *Studies in the Linguistic Sciences*, Vol. 24.2, Fall 1994.
47. Mari Broman Olsen. Telicity and English Verb Classes and Alternations: An Overview. Umiacs tr 96-15, cs tr 3607, University of Maryland, College Park, MD, 1996.
48. Martha Palmer and Zhibao Wu. Verb Semantics for English-Chinese Translation. *Machine Translation*, 9(4), 1995.
49. Terence Parsons. *Events in the Semantics of English.* The MIT Press, Cambridge, MA, 1990.
50. Steven Pinker. *Learnability and Cognition.* The MIT Press, Cambridge, MA, 1989.
51. Stanley Porter. Verbal Aspect in the Greek of the New Testament, with Reference to Tense and Mood. In D.A. Carson, editor, *Studies in Biblical Greek, Vol. 1.* Peter Lang, New York, 1989.
52. 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.

53. James Pustejovsky. The Syntax of Event Structure. In Beth Levin and Steven Pinker, editors, *Lexical and Conceptual Semantics*. Blackwell, Cambridge, MA, 1991.
54. James Pustejovsky. *The Generative Lexicon*. The MIT Press, Cambridge, MA, 1995.
55. J.J. Quantz and B. Schmitz. Knowledge-Based Disambiguation for Machine Translation. *International Journal of Minds and Machines, Special Issue on Knowledge Representation for Natural Language*, 4:39–57, 1994.
56. Roger C. Schank. *Conceptual Information Processing*. Elsevier Science Publishers, Amsterdam, 1975.
57. Leonard Schubert. Semantic Nets are in the Eye of the Beholder. In J. Sowa, editor, *Principles of Semantic Networks*. Morgan Kaufmann, San Mateo, CA, 1991.
58. Dan I. Slobin. Typology and Rhetoric: Verbs of Motion in English and Spanish. In M. Shibatani and S. A. Thompson, editors, *Grammatical Constructions: Their Form and Meaning*. Oxford University Press, Oxford, to appear.
59. Dan I. Slobin and Aura Bocaz. Learning to Talk Aobut Movement Through Time and Space: The Development of Narrative Abilities in Spanish and English. *Lenguas Modernas*, 15:5–24, 1988.
60. Carlota Smith. *The Parameter of Aspect*. Kluwer, Dordrecht, 1989.
61. J. Sowa. Toward the Expressive Power of Natural Language. In J. Sowa, editor, *Principles of Semantic Networks*, pages 157–189. Morgan Kaufmann, San Mateo, CA, 1991.
62. Leonard Talmy. Semantics and Syntax of Motion. In J.P. Kimball, editor, *Syntax and Semantics 4*, pages 181–238. Academic Press, New York, NY, 1975.
63. Leonard 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.
64. Leonard Talmy. Path to Realization: A Typology of Event Conflation. Technical Report Buffalo Papers in Linguistics, Vol. 91-01, State University of New York, Buffalo, NY, 1991.
65. Carol Tenny. The Aspectual Interface Hypothesis. In I.A. Sag and A. Szabolcsi, editors, *Lexical Matters*, pages 1–27 (An earlier version appeared in NELS 18:490– 508.). CSLI Publications, Stanford, CA, 1992.
66. Carol Tenny. *Aspectual Roles and the Syntax-Semantics Interface*. Kluwer, Dordrecht, 1994.
67. Carol Tenny. Review of Verkuyl's *A Theory of Aspectuality*. *Language*, 72:121–126, 1996.
68. Stephen Ullmann. *Semantics: An Introduction to the Science of Meaning*. Harper and Row, New York, NY, 1962.
69. Angeliek Van Hout. Projection Based on Event Structure. In P. Coopmans, Martin Everaert, , and Jane Grimshaw, editors, *Lexical Specification and Lexical Insertion*. Lawrence Erlbaum Associates, Hillsdale, NJ, To appear.
70. Robert D. Van Valin, Jr. A Synopsis of Role and Reference Grammar. In Robert D. Van Valin, Jr., editor, *Advances in Role and Reference Grammar*, pages 1–164. John Benjamins, Amsterdam, 1993.
71. Zeno Vendler. Verbs and Times. *Philosophical Review*, 56:143–160, 1957. Reprinted in *Linguistics in Philosophy*, 1967, pages 97–121, Cornell University Press, Ithaca, NY.
72. Henk Verkuyl. *On the Compositional Nature of the Aspects. Foundations of Language Supplementary Series 15*. Reidel, Dordrecht, 1972.
73. Henk Verkuyl. *A Theory of Aspectuality: The Interaction Between Temporal and Atemporal Structure*. Cambridge University Press, Cambridge and New York, 1993.

74. 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.
75. Uriel Weinreich. Webster's Third: A Critique of its Semantics. *International Journal of American Linguistics*, 30:405–409, 1964.

Address for correspondence: {bonnie,mari}@umiacs.umd.edu