

Lexical Factorization and Syntactic Behavior

JAMES PUSTEJOVSKY, *Brandeis University*
ARAVIND JOSHI, *University of Pennsylvania*

Abstract

In this paper, we examine the correlation between lexical semantics and the syntactic realization of the different components of a word's meaning in natural language. More specifically, we will explore the effect that *lexical factorization* in verb semantics has on the suppression or expression of semantic features within the sentence. Factorization was a common analytic tool employed in early generative linguistic approaches to lexical decomposition, and continues to play a role in contemporary semantics, in various guises and modified forms. Building on the unpublished analysis of *verbs of seeing* in Joshi (1972), we argue here that the significance of lexical factorization is twofold: first, current models of verb meaning owe much of their insight to factor-based theories of meaning; secondly, the factorization properties of a lexical item appear to influence, both directly and indirectly, the possible syntactic expressibility of arguments and adjuncts in sentence composition. We argue that this information can be used to compute what we call the *factor expression likelihood (FEL)* associated with a verb in a sentence. This is the likelihood that the overt syntactic expression of a factor will cooccur with the verb. This has consequences for the compositional mechanisms responsible for computing the meaning of the sentence, as well as significance in the creation of computational models attempting to capture linguistic behavior over large corpora.

1 Introduction

There is a long and established tradition in linguistics of the application of component-based decomposition for the analysis of lexical meaning. Early work in generative grammar incorporated both the feature-based

analysis promoted by Katz and Fodor (1963) and others (Katz and Postal, 1964, Chomsky, 1965), as well as the adoption of more structured lexical templates as put forth by Fillmore (1968b), Leech (1970), Gruber (1976), and Lakoff (1970), to the study of verb classes and the syntactic and semantic variations associated with class membership.

In this paper, we reevaluate the significance of lexical decomposition strategies as they were employed in early generative linguistic studies of verb meaning and syntactic variability, an area now often referred to as the *lexicon-syntax interface*, and consider their influence on contemporary treatments of argument selection and polysemy. We look at the general strategy of *lexical factorization* as a method for characterizing variations in meaning within a semantic class, focusing in particular on an unpublished 1972 manuscript of Aravind Joshi's, "Factorization of Verbs: An Analysis of Verbs of Seeing". This work adheres to a style of lexical factorization that was common in the 1970s, particularly with linguists exploring the impact of semantic factors on syntactic behavior or psychological plausibility. For example, Joshi's analysis follows a scheme introduced by Miller (1973) to analyze the conceptual relations between motion verbs, as a way of modeling lexical memory. Miller and other early researchers opened up new avenues for exploring both the psychological and semantic modeling of linguistic representations for lexical forms (Miller and Johnson-Laird (1976), Miller and Fellbaum (1991)).

We argue that the factorization methodology, as adopted in Joshi's paper and by other earlier generative linguists, is relevant to the current discussion on mapping to syntax from lexical forms for several reasons. First, componential analysis and factorization, as carried out by both Jackendoff (1972) and Miller (1973) – and exemplified in Joshi's analysis as presented here – remain important aspects of the analytic toolkit for modeling word meaning in linguistics. In fact, much of the work on semantic type systems and typed feature structures developed in part out of this work. The selectional behavior of predicates, on this view, is a reflection of the types associated with the verb's arguments, and the specificity of an argument's type reflects the meaning of the predicate in a direct manner. We will show how the modern day analyses of polysemy and argument selection in linguistics are related to these earlier models of lexical factorization, illustrating how Joshi's lexical decomposition can be re-interpreted in terms of contemporary semantic models.

We then explore the relationship between lexical factorization and the expected syntactic expressibility of semantic variables within the sentence. Joshi's (1972) analysis, along with Miller's (1972) work and

Fillmore’s early analyses in terms of case frames (Fillmore, 1968a,b) have an additional relevance and topicality for current computational approaches to language modeling and disambiguation. By analyzing lexical forms in terms of a restricted inventory of semantic features, where a given sentential frame has a fixed number of semantic parameters available to it, one is able to identify the possible factors within a sentence as being either lexically encoded or overtly expressed in the syntax.

One obvious consequence of this is that a specific lexical factorization for a verb’s word sense will influence the syntactic expressibility of these and other factors. Such an association is related to the standard information theoretic metrics used in computational models for lexical disambiguation or computing the likelihood of phrasal attachment. In fact, one can define the *factor expression likelihood* for a verb sense, based on its constituent semantic factors. This is the likelihood that the overt syntactic expression of a factor will cooccur with the verb. This has consequences for the compositional mechanisms responsible for computing the meaning of the sentence, as well as computational significance in the creation of language models attempting to capture linguistic behavior over large datasets. This work can be seen as a logical extension of the work begun in Pustejovsky and Rumshisky (2008), Pustejovsky and Jezek (2008), exploring the link between theory and distributional properties of linguistic data.

2 Factorization of Lexical Meaning

Before we explore the factorization model from Joshi (1972) in detail, it is worth briefly reviewing the basic motivation behind theories of lexical decomposition within linguistics. One of the major concerns in both early generative and non-generative (distributional) linguistic research was the question of how verb meanings relate to various syntactic forms within the sentence; that is, linking lexical-semantic forms to syntax (Chomsky, 1965, Garvin, 1967). Because of the general assumption within generative semantics that the deep structure of a sentence completely determined its meaning, the structure of lexical representation was an important research concern. Transformations could add, delete, or edit structures, but not change the semantic content conveyed by the resulting expression, i.e., the surface structure.¹ Hence, the proper

¹What is interesting about the work on decomposition from the late 1960s and the 1970s is the overall goal of the project, rather than the specific implementation that was seen in how semantic factors were projected or mapped to syntax by mechanisms within the grammar. Rules of Predicate raising, collapsing, etc. were proposed and argued about, but the undeniable effect of surface interpretation of

lexical encoding of semantic features was of critical concern (Lakoff, 1971a).² The features comprising the different aspects of semantics for lexical items were thought of as *components*, *primitives*, or *factors*, of meaning Pustejovsky (2015). From work in generative semantics (Lakoff, 1970, McCawley, 1968), through case grammar (Fillmore, 1965), Miller’s 1972 work, to early interpretive semantics (Jackendoff, 1972, 1983) some common themes emerge concerning the decompositional aspects of word meaning. We will focus on verb meaning, since the subject of Joshi’s article is on a subclass of perception predicates, the verbs of seeing. The list below is not exhaustive, but does cover many of the factors contributing to the meaning of a predicate:

- (1) a. Specific attributes of the arguments of the verb;
- b. Change of state of an individual (e.g., location, size, constitution, or other relation);
- c. The coming into / going out of being of an individual;
- d. Causation and encoding of agency;
- e. Specification of manner and means of an activity;
- f. Temporal or spatial constraints on the event;
- g. Intentionality of the actor;
- h. Specification of an instrument involved;
- i. Mention of the psychological state of the participants;
- j. Determination of the medium of the event.

Most of these have been considered as component factors in decompositional models within linguistics by Dowty, Lakoff, Jackendoff to name just a few. But how are these factors encoded in the representation of lexical meaning? For the purpose of discussion, we will abstract away from the differences between lexical semantic models, in order to better appreciate the role factorization has played in the development of current theories of semantic selection.

In order to formally distinguish the various techniques of decomposition, we first identify a predicate expression as being composed of both a *argument list* and a *body*. This is schematically illustrated in (2) below.³

$$(2) \overbrace{\lambda x_i}^{Args} \overbrace{[\mathcal{P}]}^{Body}$$

syntactic forms made many of these forms of lexical substitution models difficult to maintain.

²This becomes less of a motivation for lexical design in Jackendoff (1972), where, following Chomsky (1970), surface structure is available as a level for semantic interpretation.

³We follow Pustejovsky (2006) in this discussion.

We can think of *body* as denoting the content of a predicative term, regardless of its form or complexity. Likewise, *Args* denotes the list of parameters that participate in the expression within the body. By examining how this simple predicative template is modified, we can identify four different strategies for defining the semantics of a predicate. These are listed in (3) below, where we illustrate the differences in how verb meanings are defined.

- (3) a. ATOMIC PREDICATION: The verb meaning is represented as a primitive itself in the semantics, hence this approach is non-decompositional. The semantic parameters match the way arguments to the verb are expressed in the syntax:

$$\lambda x_n \dots \lambda x_1 [\mathcal{P}(x_1, \dots, x_n)] \Rightarrow \text{Verb}(\text{Arg}_1, \dots, \text{Arg}_n)$$

- b. PREDICATIVE FACTORIZATION: The verb meaning is represented by a complex expression of subpredicates, $\mathcal{P}_1, \dots, \mathcal{P}_k$, each bound by combinations of the semantic parameters. Special rules of lexical insertion and predicate collapsing ensure that the appropriate verbal form is expressed in syntax (where x_i and x_j are instances of x):

$$\lambda x_n \dots \lambda x_1 [\mathcal{P}_1(x_i), \dots, \mathcal{P}_k(x_j)] \Rightarrow \text{Verb}(\text{Arg}_1, \dots, \text{Arg}_n)$$

- c. PARAMETRIC ENRICHMENT: The verb meaning is represented as a primitive (predicative constant) in the semantics, \mathcal{P} , but additional parameters, p_i , are added for interpretation in the model, that are not overtly realized in the syntax.

$$\lambda p_m \dots \lambda p_1 \lambda x_n \dots \lambda x_1 [\mathcal{P}(p_1, \dots, p_m, x_1, \dots, x_n)] \Rightarrow \text{Verb}(\text{Arg}_1, \dots, \text{Arg}_n)$$

- d. STRUCTURED FACTORIZATION: The verb meaning is represented by a complex expression of subpredicates, $\mathcal{P}_1, \dots, \mathcal{P}_k$, while also adding additional parameters, binding into the subpredicates (where x_i and x_j are instances of x , and p_k and p_l are instances of p):

$$\lambda p_m \dots \lambda p_1 \lambda x_n \dots \lambda x_1 [\mathcal{P}_1(p_i, x_k), \dots, \mathcal{P}_k(p_l, x_j)] \Rightarrow \text{Verb}(\text{Arg}_1, \dots, \text{Arg}_n)$$

Early approaches to modeling language meaning with formal logic, such as Montague (1970), as well as Fodor's more recent position Fodor (1998, 2000), are non-decompositional, and can be identified with the *atomic predication* approach in (3a). On this view, if inferences are to be made between members of a language's lexis, then they are accomplished by meaning postulates or explicit axioms. Hence, related meanings of a word are encoded as separate lexical entries; for verbs,

even variations in subcategorization frames must be encoded as distinct entries.

Consider the impact this approach has for the interpretation of a verb’s meaning, for example, the perception verb, *see*. Assuming the basic veridicality of perception reports (cf. (Hintikka, 1969)), we can infer the factivity of what we see.⁴ Hence, from (4a) we can infer (4b). Further, from our understanding of vision as a kind of perception, we can also infer (4c) from (4a).

- (4) a Mary saw John fall.
 b. \models John fell.
 c. \models Mary perceived John fall by sight.

These inferences are not part of the verb’s meaning, but rather are accomplished with the help of rules or axioms that accompany the atomic meaning of lexical items.

Now consider the same verb analyzed in terms of lexical factors, following the strategy of structured factorization, described in (3d).⁵ One of the motivations for structured predicative decomposition is the ability to read inferences such as those above directly from the semantic form of the sentence. That is, by decomposing *see* in (5) below, into the three lexical factors SIGHT, PERCEIVE, and BY-MEANS, the inferences in (4b) and (4c) are directly recoverable (where *y* is typically typed as THING or EVENT, as in (4a)).

- (5) x PERCEIVE y BY-MEANS SIGHT

In the discussion below, we argue that the tradition of structured factorization opened new avenues for studying semantic associations within the lexicon, and paved the way for systematic models of argument selection. In the next section, however, we first review the contents of Joshi’s 1972 manuscript, “Factorization of Verbs: An Analysis of Verbs of Seeing”, and examine how factors are deployed to model verb classes and lexical ambiguity within classes.

3 Joshi’s “Verbs of Seeing”

A major motivation for lexical factorization and decompositional approaches to word meaning was the belief that lexical semantic templates for words largely determine their syntactic behavior in the language.

⁴For the current discussion, we ignore the problem of visual illusions (van der Does and Van Lambalgen, 2000).

⁵For the present discussion, we will ignore simple predicative approaches as pursued by Katz and Fodor (1963), and parametric strategies, as utilized in theorem proving and AI (Hobbs, 1995). See, however, Pustejovsky (2006) for more extensive treatment of these approaches.

Joshi's unpublished article, "Factorization of Verbs: An Analysis of Verbs of Seeing" (henceforth VOS), is a study of how this is realized within the class of visual perception verbs. He gives the criteria for inclusion in this general class as follows:

A seeing verb involves visual perception in many different ways: (a) The subject of the verb sees something or the object of the verb becomes visible (e.g. see: X saw Y); (b) The subject directs sight in order to see (e.g. look: X looked at Y); (c) The subject became visible to someone (e.g. appear: X appeared); (d) The basic predicates of see, look, and appear are 'modulated' (modified) by other predicates such as cause, not, momentary, continual, etc. yielding a variety of seeing verbs such as hide, expose, glance, gaze, etc. (p. 22, Joshi (1972))

The modulating predicates mentioned above are what we refer to as *factors* of meaning: decompositional elements similar in some respects to the predicates used in generative semantics (Lakoff, 1970, McCawley, 1968, Ross, 1970, Green, 1973). The semantic variability within this class is accounted for with combinations of operator-operand expressions, a technique employed with regularity in early generative analyses

Adopting the strategy used in Miller (1972), Joshi first identifies the core element of meaning that is shared by all verbs in the class of visual perception predicates. This is shown in (6) along with its representation in terms of factors.

- (6) a. x perceives a thing y by sight
 b. $see_0(x, y) =_{df} \text{BY-MEANS}(\text{PERCEIVE}(x, \text{THING}(y)), \text{SIGHT})$

We will call this the *base* element for the class. The base in (6b) is composed of a primitive predicate, PERCEIVE and a number of factors wrapping and enriching this basic predicative unit. It provides the interpretation for the most common sense for the English verb *see*, and is used to construct the representation for the family of visual perception verbs by composition with additional factorization.

Joshi also follows Miller in identifying the most important types of evidence for factorization as: a. Syntactic-semantic evidence; b. Semantic evidence; and c. Psychological evidence. Unlike the analyses in generative semantics, however, Miller and Joshi eschew an interpretation of decomposition as providing a complete determination of the meaning for a lexical item. Hence, both the base element of a verb class and the subsequent factorizations of the other verbs are to be taken as *incomplete definitions*.

Joshi's paper analyzes around 60 verbs of seeing, given here: *appear, flash, scrutinize, behold, flaunt, secrete, blind, gaze, screen, blur, glance, see, conceal, glare, search, cover, glimpse, seek, darken, hide,*

show, descry, inspect, sight, disclose, look, spot, dazzle, mask, stare, display, notice, surface, disappear, obscure, survey, discern, observe, uncover, emerge, overlook, unmask, espy, peek, vanish, examine, peep, view, expose, peer, watch, exhibit, reappear, witness, eye, regard, fade, resemble, fade-in, reveal, fade-out, scan.

There are eleven basic subtypes that are identified within this class, each represented by a unique lexical factorization. Further, within each subclass, the various word senses are themselves distinguished by uniquely factorized expressions. There are two immediate consequences of organizing lexical items in this fashion.

- (7) a. *Word sense disambiguation*: the presence of lexical factors influences the likelihood of factor expression in the sentence the lexical item appears in.
 b. *General factor realization over an entire frame* (semantic expression for a sentence): Independent of a lexical disambiguation decision, it is possible to influence the determination of how to interpret possibly ambiguous (syntactically through attachment or semantically through denotation) adjunct phrases and modifiers.

3.1 See

Let us consider first the verb most directly associated with the base element of the class, *see*₀: this is obviously the verb *see*, as used in (8a) below.

- (8) a. Mary saw a snake.
 BY-MEANS(PERCEIVE(x,THING(y)), SIGHT)
 b. Mary saw a snake crawl into her backpack.
 BY-MEANS(PERCEIVE(x,EVENT(y)), SIGHT)
 c. Mary saw that John had visited.
 BY-MEANS(PERCEIVE(x,FACT(y)), SIGHT)

The uses in (8b) and (8c) involve minor modifications to the base, where the type of *y*, THING, has been replaced by EVENT and FACT, respectively.

3.2 Look

Now, let us examine the subclasses derived from the base through additional semantic factorization. Consider the meaning associated with the verb *look*, as used in (9) below.

- (9) a. John looked at the tree.
 b. John looked into the cave.
 c. John looked towards the tree.

The subcategorization for *look* (in this sense) allows only a prepositional object. This class subsumes the base, see_0 , while wrapping it with an INTEND operator, $INTEND(x, see_0)$, as in to perform intentionally:

(10) $INTEND(x, BY-MEANS(PERCEIVE(x, THING(y)), SIGHT))$

Unlike the **see**-class, Joshi (1972) claims that no eventive or factive reading is possible with this class, as in (11). Contrary to this view, sentence (12) does seem acceptable, however.

(11) a. ?Mary looked at the snake crawl into her backpack.

b. ?Mary looked at John's having visited.

(12) Mary looked at the snake crawling into her backpack.

3.3 Glance

This class adds a temporal factor, **MOMENTARY**, to the enriched form encoding the **look**-class, where it modifies the duration of the perception act, $INTEND(x, MOMENTARY(see_0))$:

(13) $INTEND(x, BY-MEANS(MOMENTARY(PERCEIVE(x, THING(y))), SIGHT))$

The verb *glance* has this interpretation as used below.

(14) a. Mary glanced at the tree.

b. John glanced at his watch.

As with the **look**-class, the object must be of type **THING**.

3.4 Notice

This is similar to the **see**-class, except that it adds a presupposition of novelty, which Joshi characterizes as factive in nature, i.e., **FACT** (although not all facts are novel). Consider the following pair:

(15) a. John saw a missing book.

b. John noticed a missing book.

Sentence (15a) refers to a direct perception event of John seeing something (a book that had gone missing), while (15b) refers to a factive, that there is a book missing. Importantly, John does not actually perceive the book in this case. So, this class can be characterized as:

(16) $BY-MEANS(PERCEIVE(x, FACT(y)), SIGHT)$

3.5 Examine

This verb embeds the structure of the **look**-class and adds the factor of a **PURPOSE**.

(17) a. Mary examined the painting for a signature.

b. John examined the car for damage.

- (18) IN-ORDER-TO(INTEND(x,BY-MEANS
(PERCEIVE(x,THING(y)), SIGHT)), PURPOSE)

Because there is an encoded purpose for the activity, sentences without explicit clarification of what that purpose is, such as (19), will presuppose such information; similarly, sentences that either contradict this encoding or supply no new information will be infelicitous, as in (20).

- (19) a. Mary examined her face.
b. John examined his steak.
- (20) a. ?Mary examined the painting accidentally.
b. ?John examined the car on purpose.

3.6 Look for

In some ways, this verb specializes the activity designated as the purpose, mentioned in the previous class; namely, the purpose is FIND. The encoding is more complex, however, since the thing is not perceived until (if) it is actually found. Hence, the perceptual activity that the agent is engaged in involves objects that are not the intended object.

- (21) IN-ORDER-TO(INTEND(x,BY-MEANS
(PERCEIVE(x,THING(z)), SIGHT)), FIND(x,THING(y)))
- (22) a. John looked for his keys for hours.
b. Mary looked for a parking spot.

3.7 Appear

The next class employs an inchoative factor (what Joshi refers to as BEGIN), usually expressed with the BECOME operator in other factorization analyses from the period (cf. Lakoff, 1970, McCawley, 1968, Jackendoff, 1972). Consider the sentences in (23) below.

- (23) a. John appeared at the door.
b. A large ship appeared on the horizon.

Joshi analyzes *appear* as an inchoative of the possibility of being seen:

- (24) BECOME(POSSIBLE(BY-MEANS (PERCEIVE(x,THING(y)), SIGHT)))

3.8 Fade

Although Joshi does not propose a distinct class for the antonym of *appear*, the factorization of *disappear* is straightforward, as used in sentences such as (25).

- (25) a. The book disappeared from the library.
b. The boat disappeared behind the island.

- (26) BECOME(NOT (POSSIBLE(BY-MEANS
(PERCEIVE(x,THING(y)), SIGHT))))

A verb that Joshi does examine and which is semantically related to *disappear* is *fade*, as used in (27) below.

- (27) a. The signature has faded.
b. Her tan faded quickly.

Joshi introduces the factor GRADUAL to indicate the incremental or graded nature of the change.

- (28) GRADUAL(BECOME(NOT (POSSIBLE
(BY-MEANS (PERCEIVE(x,THING(y)),SIGHT))))))

This can be compared to later classifications of verbs as involving incremental change of state (Tenny, 1994, Dowty, 1991, Krifka, 1989).

3.9 Expose

The next two classes are causal inverses, building on top of the previous two classes, **appear** and **fade**. First, consider the verb *expose*, as in *reveal*.

- (29) a. John exposed himself to the enemy.
b. The erosion exposed the hard volcanic core of these domes.

- (30) CAUSE(z,BECOME(POSSIBLE(BY-MEANS
(PERCEIVE(x,THING(y)), SIGHT))))

Second, consider its inverse, *hide*:

3.10 Hide

- (31) a. John hid the jewels under the table.
b. Terrorists were hiding inside the plane.

- (32) CAUSE(z,BECOME(NOT (POSSIBLE(BY-MEANS
(PERCEIVE(x,THING(y)), SIGHT))))

3.11 Blind

- (33) a. The storm blinded me.
b. The paparazzi flashbulbs blinded her steps.

Although Joshi originally proposed the factorization for *blind* as given in (34a), it can also be expressed as (34b), given the inventory of factors available.

- (34) a. CAUSE(ACTION,BECOME(NOT (POSSIBLE(BY-MEANS
(PERCEIVE(x,THING(y)), SIGHT))))
b. BY-MEANS(CAUSE(z,BECOME(NOT (POSSIBLE(BY-MEANS
(PERCEIVE(x,THING(y)), SIGHT))))),ACTION)

This distinction, however, does not bear on subsequent discussion.

4 Factors and Argument Selection

Having reviewed the basic factorization analysis of perception predicates as presented in Joshi (1972), we turn to how aspects of this theory can be viewed within current approaches to argument selection. We will discuss the relation between selection and the elements that are assumed as part of the type inventory of the compositional semantic system.

In almost all treatments of lexical decomposition within generative linguistics from the 1960s and 1970s, the arguments to the predicates are taken as variables in a standard first-order domain. Hence, any restrictions to the predication are handled as identifiers or classifier predicates. As mentioned above, a richer system of types works to effectively introduce the test of “possible satisfaction” of an argument to a predicate by restricting the kinds of arguments that can combine with a predicate.

Let us walk through how we can interpret the verb factorizations encountered in the previous section within a richer typing environment. Consider the expression for a two-place predicate, \mathcal{P} , which consists of the lexical factors $\mathcal{P}_1, \dots, \mathcal{P}_k$. The variables are typed as individuals, i.e., e , and the entire expression is therefore a typical first-order relation, typed as $e \rightarrow (e \rightarrow t)$.

$$(35) \quad \overbrace{\lambda x_2 \lambda x_1}^{Args} \left[\overbrace{\mathcal{P}_1, \dots, \mathcal{P}_k}^{Body} \right]$$

This is the verb factorization as used in Joshi’s analysis. Now let us substitute the factorization for the class *see* into the body of the expression.

$$(36) \quad \overbrace{\lambda y \lambda x}^{Args} \left[\overbrace{\text{BY-MEANS}(\text{perceive}(\text{human}(x), \text{thing}(y)), \text{sight})}^{Body} \right]$$

How does this correspond to selectional models that assume richer systems of types, such as HPSG, LFG, CCG, or GL? Joshi (1972) identified five distinct senses for the verb *see*.

$$(37) \quad \begin{aligned} &\text{a. } \lambda y \lambda x [\text{BY-MEANS}(\text{perceive}(\text{human}(x), \text{thing}(y)), \text{sight})] \\ &\text{b. } \lambda y \lambda x [\text{BY-MEANS}(\text{perceive}(\text{human}(x), \text{event}(y)), \text{sight})] \\ &\text{c. } \lambda y \lambda x [\text{BY-MEANS}(\text{perceive}(\text{human}(x), \text{fact}(y)), \text{sight})] \\ &\text{d. } \lambda y \lambda x [\text{BY-MEANS}(\text{ensure}(\text{human}(x), \text{fact}(y)), \text{sight})] \\ &\text{e. } \lambda y \lambda x [\diamond [\text{BY-MEANS}(\text{perceive}(\text{human}(x), \text{fact}(y)), \text{sight})]] \end{aligned}$$

The first three are variants, differing only in the typing restrictions

associated with a specific argument (what Joshi calls a “classifier” or “identifier” predicate). A richer typing structure for the arguments can identify such predicates in the body of the expression, and restrict the application of the function to arguments satisfying that type.

$$(38) \lambda y \lambda x [\mathcal{P}_1, \dots \overbrace{\text{human}(x_1)}^{\tau}, \dots \overbrace{\text{thing}(x_2)}^{\sigma}, \dots, \mathcal{P}_k]$$

Notice that by identifying the sortal restriction factors from the predicate body as types, we can generalize the core semantics of the expression. That is, we take the identifier predicates associated with each argument and *reify* them as type restrictions on the λ -expression, i.e., as the types τ (HUMAN) and σ (THING).

$$(39) \lambda y : \text{THING } \lambda x : \text{HUMAN} [\mathcal{P}_1, \dots, \mathcal{P}_k - \{\sigma, \tau\}]$$

The result of this is to take us from the factorization in (40a) below, to the typed expression in (40b).

$$(40) \text{ a. } \lambda y \lambda x [\text{BY-MEANS}(\text{perceive}(\text{human}(x), \text{thing}(y)), \text{sight})] \\ \text{ b. } \lambda y : \text{THING } \lambda x : \text{HUMAN} [\text{BY}(\text{perceive}(x, y), \text{sight})]$$

Notice that three of the representations corresponding to distinct senses for the verb *see* can be encoded as sharing the same base lexical form, if “classifier” factors are expressed as typing distinctions:

$$(41) \lambda y : \sigma \lambda x : \text{HUMAN} [\text{BY-MEANS}(\text{perceive}(x, y), \text{sight})] \\ \text{ where } \sigma \in \{\text{THING, EVENT, FACT}\}.$$

This allows us to express the logical polysemy of the first three senses more clearly, while at the same time, differentiating these from the derived richer senses found in (37d-e).

When semantic factors are lexically encoded, they behave in many respects like arguments. They may even license selection. This has been discussed in terms of the different kinds of arguments that a verb may select. Within Generative Lexicon, for example, there is a distinction made between conventional arguments (obligatory and default) and *shadow* arguments, those that can appear only if the syntactic realization of the factor provides new information Pustejovsky (1995). For example, consider the phenomenon of incorporation seen in sentence (42) and (43) below.

- (42) a. Harry elbowed me.
 b. !Harry elbowed me with his elbow.
 c. Harry elbowed me with his arthritic elbow.

- (43) a. Mary buttered the bread.
 b. !Mary buttered the bread with butter.
 c. Mary buttered the bread with creamy, unsalted butter.

What these examples illustrate is the relationship between information that has been incorporated into the verb, and how it can be used to license the expression of arguments, under certain conditions. On this analysis, both verbs *elbow* and *butter* carry a *shadow* argument, which can only be overtly expressed if it is “more informative” ((Pustejovsky, 1995).

- (44) **Shadow Argument Licensing:** A shadow argument, x_s , of type τ , to a verb, v , can be expressed syntactically by a phrase, X , within a sentence, only if the denotation of X is more informative than τ (i.e., it is a subtype or elaboration).

How does this relate to the factorization analysis as employed by Joshi in the previous section? Following the decomposition strategy developed in Joshi’s paper and Miller and Johnson-Laird (1976), the verb *butter* can be glossed as follows, where the semantic factor of BUTTER is incorporated into the verb’s meaning, building on a base representation of the relation SPREAD. We will adopt the distinction between obligatory (x) and shadow (x_s) arguments introduced above.

- (45) **butter** =_{df} $\lambda z_s : \text{BUTTER} \lambda y : \text{PHYS} \lambda x : \text{HUMAN}[\text{spread}(x, y, z)]$

The behavior of the sentences in (43) suggest that lexical factorization can block the syntactic realization of a shadow argument, (43b), unless it is more informative, as in (43c).

5 Some Thoughts on Factorization in Syntax

The relevance of factorization for lexical choice and syntactic realization comes from some obvious properties of how factors encode a word’s meaning and how this impacts the subsequent syntactic behavior of the word in a sentential context; that is, there is a correlation between the information encoded in a verb v , and the expression of factors for a particular situational frame in which it is used. It has long been recognized that correlations between lexical items and the properties of selection can be linked to patterns of linguistic behavior in data (Harris (1957), Firth (1961)), and this discussion relates directly to work done in corpus and distributional linguistics, starting with Church and Hanks (1990), Resnik (1993), Hindle and Rooth (1993), up to the present.

Recall from previous discussion that, within a situational frame, a semantic factor, f_i , can be either *lexically encoded*, *syntactically realized*, or not expressed at all. In fact, it might be expected that lexical and syntactic factorization is inversely correlated in an utterance: that is, if a verb already incorporates a semantic factor, f_i , then it is less likely to be expressed syntactically. Likewise, if the verb does not incorporate

such a factor, it might be more common to see it appear in the syntax. We refer to this as an informal statement of the *factor expression likelihood (FEL)*, which we characterize more formally below.

In order to measure the likelihood of a specific semantic factor being expressed syntactically, we need some characterization of the syntactic constructions associated with the factor. Let us adopt the following thesis.

- (46) **The Factor Expression Thesis:** Most semantic components (factors) can be identified with a fixed set of canonical syntactic realization strategies.

This is essentially a generalization of semantic mapping strategies that have been proposed since Fillmore (1968b), McCawley (1968), and Jackendoff (1972). To illustrate how this works, we see how named semantic relations (e.g., thematic or case roles) can systematically project to syntax.

- (47) a. INSTRUMENT: “*with NP*”, “*through NP*”
 Mary saw the star with a telescope / through a telescope.
 b. LOCATION: “*through/in NP*”
 Mary looked for the watch in the park.
 c. MEDIUM: “*through/in NP*”
 The bird flew in the air.

Now consider some of the factors encoded in the *see*-class by Joshi’s analysis. The examples in (48) illustrate both how a factor, f_i , is syntactically realized and how it is lexically encoded in a semantically incorporating verb.

- (48) a. BY-MEANS: “*by V-ing*”
syn: Mary perceived the star by seeing it.
lex: Mary saw the star.
 b. CAUSE: “*make/cause/let*”
syn: The sun made John unable to see.
lex: The sun blinded him.
 c. IN-ORDER-TO: “*to VP*”, “*in order to VP*”
syn: John looked at the painting closely in order to check its authenticity.
lex: John examined the painting for its authenticity.

The factor expression likelihood (FEL) associated with semantically related verb pairs should reflect the decomposition associated with each lexical item. For example, consider the English verbs *kill* and *murder*. As argued in Lakoff (1971b), Katz (1972) and elsewhere, both these verbs entail an agent causing someone’s death, but the latter verb en-

codes that it is performed intentionally; that is, it has the additional semantic factor of INTEND. Assuming the canonical syntactic expression for this factor is an adverbial phrase with specific lexicalizations (e.g., *intentionally*, *on purpose*), then one would expect that this factor would syntactically express for *kill* but not for *murder*, and this seems intuitively to be the case. Given such observations, the correlation between lexicalized factors and syntactic expression might be more explicitly formulated as follows.

- (49) **Factor Uniqueness Hypothesis (version 1)**: Each semantic component (factor) contributing to the meaning of an utterance is *uniquely expressed* in composition. That is, if a factor, f_i , is lexically encoded by a word, w , then it cannot be independently expressed syntactically; and if f_i is expressed syntactically by a phrase, X , it may not be lexically encoded by any word not part of X .

Given what we have said, however, we have no means of testing whether such a claim is true: namely, without a more formal characterization of the likelihood of a factor being expressed syntactically or not, we cannot measure the dependency between lexical incorporation and syntactic projection.

In order to better understand the relationship between incorporation and expression, let us try to give more substance to the notion of FEL. Intuitively, this is the likelihood that there will be an overt syntactic expression for a factor, f_i , given a specific verb, v . For all the factors, \mathbf{f} , associated with a situational frame, we are interested in the probability of each of these factors, f_i , occurring in an utterance, conditioned on v being selected; i.e., $P(\mathbf{f}|v)$.⁶

But this is not entirely correct, since we want to take into consideration the prior probability of all the factors being expressed, independently of the verb, i.e., $P(\mathbf{f})$. This we can do by normalizing the conditional, $P(\mathbf{f}|v)$, which is equivalent to $\frac{P(\mathbf{f},v)}{P(v)}$, by $P(\mathbf{f})$, giving us:

$$(50) \frac{P(\mathbf{f},v)}{P(\mathbf{f})P(v)}$$

If we focus on each specific factor value, f_i , rather than the random variable, \mathbf{f} , then this is, in fact, the well-known concept of *pointwise mutual information* (Fano (1961)):

$$(51) \text{PMI}(f_i, v) = \log_2 \frac{P(f_i, v)}{P(f_i)P(v)}$$

⁶With this assumption, we can express the estimation of the lexical choice for a certain verb, \hat{v} , in a particular situation, with given factors, \mathbf{f} as: $\hat{v} = \arg \max_{v \in C} P(\mathbf{f}|v)P(v)$.

VERB	PMI	ADVERB_SET
kill	6.26	intentionally
-	7.98	accidentally
murder	≈ 0	intentionally
-	≈ 0	accidentally

TABLE 1 PMI results

The pointwise mutual information (PMI) compares the probability of a specific factor, f_i , and the verb, v , showing up together (jointly), with the probabilities of seeing f_i and v independently. As in Church and Hanks (1990), where it is argued “word association norms” can be measured in terms of PMI given enough linguistic data, the association between factor expression and lexical choice is also potentially captured with such a metric. While this is a relatively simple measure, and there have been many suggested improvements and alternatives to it (cf. (Merlo and Stevenson, 2001, Rumshisky, 2008, Pustejovsky et al., 2004), it will suffice to demonstrate our thesis for the remainder of this section.

Let us return to the example of the verbs *kill* and *murder* first mentioned above. We measure their relative PMI scores for two specific lexicalizations of the intentionality factor, as calculated over a large corpus of English.⁷ If there is a real correlation between a verb and an adverb of intention, then the PMI score will be significantly larger than 0 ($\gg 0$). If, on the other hand, there is no relationship of interest between them, then $\text{PMI} \approx 0$. What we find in Table 1 is that, as intuitively predicted, there appears to be a correlation between the lexical encoding of a factor, f_i , and the absence of any syntactic realization of that factor.

In other words, semantic incorporation and syntactic projection do, in fact, seem inversely correlated for factors, as stated in (49). However, when we extend this methodology to examine the factorization of the verbs of seeing, this correlation cannot be so strongly maintained, and the situation becomes much more nuanced. We turn to these problems below.

Recall the semantic factor identified to distinguish the verbs *look* and *glance*; i.e., the introduction of a temporal element, indicated by MOMENTARY.

(52) a. *look*: $\text{INTEND}(x, \text{see}_0)$:

⁷We use the enTenTen 2012 corpus, as compiled and indexed by SketchEngine. This is a corpus of close to 28 million words, and has been prepared and indexed for linguistic analytics by SketchEngine (Kilgarriff et al. (2014); <http://www.sketchengine.co.uk>).

VERB	PMI	ADVERB_SET
glance	8.5	DIRECTION: sideways, backwards, forwards, downward, skyward
-	6.42	TIME: briefly, again, occasionally, periodically, quickly, swiftly
look	12.22	DIRECTION: forward, backward, ahead, down, straight
-	≈ 0	TIME: briefly, again, occasionally, periodically, quickly, swiftly

TABLE 2 PMI results

John looked at the moon.

b. *glance*: INTEND(x , MOMENTARY(see_0))

John glanced at the moon.

While there is no indication of how long the event in (52a) was, it is strongly suggested that the event in (52b) was brief. Given that the semantic factor MOMENTARY is lexically encoded as part of *glance* and not part of *look*, we would expect that the FEL for this factor would be greater for *look* than for *glance*.

Surprisingly, what we see is exactly the opposite behavior. After performing PMI calculations over the same corpus as before, what emerges is that there is a *greater* likelihood for expression of the factor with the verb *glance* over *look*, as shown in Figure 2.

This table illustrates the average PMI values for each verb and two semantically identified adverbial classes, i.e., DIRECTION and TIME.⁸ While these data certainly disprove the Factor Uniqueness Hypothesis as given in (49), they are reminiscent of the phenomenon we encountered in the previous section with shadow argument selection.

Recall that, in certain contexts, some semantically incorporated material to a verb can license the presence of an argument or adjunct in the syntax. For example, the incorporated factors of BUTTER, LEG, and DANCE, respectively, are realized syntactically, in the sentences below.

- (53) a. Mary buttered her bread with margarine.
 b. John kicked the ball with his left leg.
 c. Mary and John danced a polka.

This follows from the shadow argument principle in (44), where, in each case, the expressed argument is more informative than the lexically encoded factor.⁹

⁸There could be a range of modifiers but adverbs lend themselves readily to this analysis.

⁹In Batiukova and Pustejovsky (2013), informativeness is used as a criterion to

Examining the PMI data in Table 2, we notice the same phenomenon: namely, the TIME factor encoded in the verb *glance* is actually acting to *select* specializations of temporal adverbials, rather than block them. This is the behavior we observed in (53). In other words, the verb *glance* can be viewed as having promoted the incorporated temporal factor as a shadow argument, as illustrated below.

$$(54) \text{ glance} =_{df} \lambda z_s:\text{TIME_MOMENT} \lambda y:\text{PHYS} \lambda x:\text{HUMAN}[\text{glance}(x, y, z)]$$

On this analysis, the temporal adverbials present in Table 2 for *glance* can be interpreted as more informative than the underlying temporal factor, in that they are specializations or elaborations of the MOMENTARY factor.

With these data, we can restate the conditions under which a factor is realized syntactically, as formulated in the Factor Uniqueness Hypothesis in (49) above as follows:

(55) **Factorization Uniqueness Hypothesis (revised version):**

Each semantic component (factor) contributing to the meaning of an utterance is *uniquely expressed* in composition. That is, if a factor, f_i , is lexically encoded by a word, w , then it cannot be independently expressed syntactically, unless it conveys new information; and if f_i is expressed syntactically by a phrase, X , it may not be lexically encoded by any word not part of X , unless X conveys new information beyond f_i .

6 Conclusion

In this brief note, we have attempted to frame the original theoretical questions motivating the Joshi (1972) analysis of the verbs of seeing in terms of the current debate concerning the lexicon-syntax interface. We have cast the analysis in terms of recent developments in lexical semantics, and then examined the consequences of this analysis with respect to distributional interpretations of syntactic behavior. We introduced the notion of *factor expression likelihood* to capture the dependency of a lexical factor on its independent expressibility in the syntax. While the results discussed here are preliminary, they suggest that a purely distributional accounting of the correlation of syntactic behavior with encoded factors must be discarded in favor of a view allowing for a more nuanced theory of semantic selection. These issues are obviously ripe for future investigations, which should cover a much wider range of verbs and their modifiers.

license adjectival modification in Adjective Noun composition.

References

- Batiukova, O. and J. Pustejovsky (2013). Informativeness constraints and compositionality. In R. Sauri, N. Calzolari, C.-R. Huang, A. Lenci, M. Monachini, and J. Pustejovsky (Eds.), *6th International Conference on Generative Approaches to the Lexicon*, pp. 92–100.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. The MIT press.
- Chomsky, N. (1970). Remarks on nominalization. In R. A. Jacobs and P. Rosenbaum (Eds.), *Readings in English Transformational Grammar*, pp. 184–221. Ginn.
- Church, K. W. and P. Hanks (1990). Word association norms, mutual information, and lexicography. *Computational linguistics* 16(1), 22–29.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language* 67, 547–619.
- Fano, R. M. (1961). *Transmission of Information: a Statistical Theory of Communications*. The MIT Press.
- Fillmore, C. J. (1968a). The case for case. In E. Bach and R. Harms (Eds.), *Universals in Linguistic Theory*. Holt, Rinehart, and Winston.
- Fillmore, C. J. (1968b). Lexical entries for verbs. *Foundations of Language* 4, 373–393.
- Firth, J. R. (1961). *Papers in Linguistics 1934-1951*. Oxford University Press.
- Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford University Press.
- Fodor, J. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. The MIT Press.
- Garvin, P. L. (1967). The automation of discovery procedure in linguistics. *Language* 43(1), pp. 172–178.
- Green, G. (1973). *Semantics and Syntactic Irregularity*. Cambridge University Press.
- Gruber, J. (1976). *Lexical structures in syntax and semantics*. North-Holland.
- Harris, Z. S. Z. (1957). Co-occurrence and transformation in linguistic structure. *Language* 33(3), 283–340.
- Hindle, D. and M. Rooth (1993). Structural ambiguity and lexical relations. *Computational Linguistics - Special Issue on Using Large Corpora* 19, 103–120.
- Hintikka, J. (1969). *Models for modalities: selected essays*. D. Reidel Publishing.
- Hobbs, J. (1995). Ontological promiscuity. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, pp. 61–69.
- Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. The MIT Press.
- Jackendoff, R. S. (1983). *Semantics and Cognition*. The MIT Press.
- Joshi, A. (1972). Factorization of verbs: An analysis of verbs of seeing.

- Katz, J. J. (1972). *Semantic theory*. Harper and Row.
- Katz, J. J. and J. A. Fodor (1963). The structure of a semantic theory. *Language* 39, 170–210.
- Katz, J. J. and P. M. Postal (1964). *An integrated theory of linguistic descriptions*. The MIT Press.
- Kilgarriff, A., V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý, and V. Suchomel (2014). The sketch engine: ten years on. *Lexicography* 1(1), 7–36.
- Krifka, M. (1989). Nominal reference, temporal constitution and quantification in event semantics. In *Semantics and contextual expression*. Dordrecht: Foris Publication.
- Lakoff, G. (1965/1970). *Irregularity in syntax*. Holt, Rinehart, and Winston.
- Lakoff, G. (1971a). On generative semantics. In D. D. Steinberg and L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics, and psychology*, pp. 232–296. Cambridge University Press.
- Lakoff, G. (1971b). Presupposition and relative well-formedness. In D. D. Steinberg and L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics, and psychology*, pp. 329–340. Cambridge University Press.
- Leech, G. (1970). *Towards a semantic description of English*. Indiana university studies in the history and theory of linguistics. Indiana University Press.
- McCawley, J. D. (1968). *Lexical insertion in a transformational grammar without deep structure*. University of Chicago Press.
- Merlo, P. and S. Stevenson (2001). Automatic verb classification based on statistical distribution of argument structure. *Computational Linguistics* 27(3), 373–408.
- Miller, G. A. (1973). English verbs of motion: A case study in semantics and lexical memory. In *Coding processes in human memory*, pp. 335–372. J.Wiley and Sons.
- Miller, G. A. and C. Fellbaum (1991). Semantic networks of English. *Cognition* 41(1), 197–229.
- Miller, G. A. and P. N. Johnson-Laird (1976). *Language and perception*. Belknap Press.
- Montague, R. (1970). Universal grammar. *Theoria* 36, 373–398. Reprinted in *Formal philosophy*, by Richard Montague, Yale University Press, 1974, pp. 222–246.
- Pustejovsky, J. (1995). *The generative lexicon*. The MIT Press.
- Pustejovsky, J. (2006). Type theory and lexical decomposition. *Journal of cognitive science* 6, 39–76.
- Pustejovsky, J. (2015). Lexical semantics. In P. Dekker and M. Aloni (Eds.), *Handbook of formal semantics*. Cambridge University Press.

- Pustejovsky, J., P. Hanks, and A. Rumshisky (2004). Automated induction of sense in context. *Proceedings of the 20th international conference on computational linguistics - COLING '04*, 924–932.
- Pustejovsky, J. and E. Jezek (2008). Semantic coercion in language: beyond distributional analysis. In A. Lenci (Ed.), *Distributional models of the lexicon in linguistics and cognitive science, special issue of Italian journal of linguistics / Rivista di linguistica*.
- Pustejovsky, J. and A. Rumshisky (2008). Between chaos and structure: Interpreting lexical data through a theoretical lens. *International Journal of Lexicography* 21(3), 337–355.
- Ross, J. R. (1970). On declarative sentences. In R. Jacobs and P. Rosenbaum (Eds.), *Readings in English transformational grammar*, pp. 272. Ginn.
- Rumshisky, A. (2008). *Verbal polysemy resolution through contextualized clustering of arguments*. Ph. D. thesis, Brandeis University.
- Tenny, C. L. (1994). *Aspectual roles and the syntax-semantics interface*. Springer.
- van der Does, J. M. and M. Van Lambalgen (2000). A logic of vision. *Linguistics and Philosophy* 23(1), 1–92.