ARTICLE IN PRESS
PLREV:116

ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect
Physics of Life Reviews ••• (••••) •••–•••

PHYSICS of LIFE reviews
www.elsevier.com/locate/plrev

Review

Embodied language, best-fit analysis, and formal compositionality

Jerome Feldman [*]

International Computer Science Institute, 1947 Center Street, Suite 600, Berkeley, CA 94704, United States

Received 26 March 2010; received in revised form 9 April 2010; accepted 10 April 2010

Communicated by L. Perlovsky

Abstract

This review brings together two fundamental, but unreconciled, aspects of human language: embodiment and compositionality.

One major scientific advance in recent decades has been *Embodiment* – the realization that scientific understanding of mind and language entails detailed modeling of the human brain and how it evolved to control a physical body in a social community.

The ability to learn and use language is one of the most characteristically human traits. Many animals signal, but only people can express and understand an essentially unbounded range of messages. The technical term for the ability of human language to support all these messages from a few dozen alphabetic symbols is *Compositionality*.

Rigor is essential for the advancement of any science, but there has been essentially no overlap between efforts to formalize language compositionality and the manifest embodiment of thought. Recent developments suggest that it is feasible to formalize the compositionality of embodied language, but that this requires a focus on conceptual composition and better understanding of contextual best-fit.

Keywords: Embodiment; Compositionality; Best-fit analysis; Language community; Construction grammar; Mental simulation

---

Contents

---

[*] Tel.: +1 510 666 2900; fax: +1 510 6662956.
E-mail address: feldman@icsi.berkeley.edu.

1571-0645/$ – see front matter
doi:10.1016/j.plrev.2010.06.006

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

---

1. Introduction

One major scientific advance in recent decades has been *Embodied Cognition* – the realization that scientific understanding of mind and intelligence entails detailed modeling of the human brain and how it evolved to control a physical body in a social community [1]. Formalization and computational modeling are essential to the science of embodied cognition, but traditional formal approaches have not proved adequate and actually constitute a barrier to progress. Until recently, formal research on language and thought was explicitly couched in disembodied terms, often under the rubric of "functionalism" [2]. This has led many investigators to define their embodied theories of mind in opposition to formal treatments. This is a disaster all around and is only slowly being corrected.

There is now a wide consensus that embodiment is an essential attribute of the human mind and language. This is an instance of the "continuity principle" of the American Pragmatists like John Dewey and William James and was also stressed by Alan Turing in his seminal 1948 paper, *Intelligent Machinery*. In discussing possible tasks for AI, Turing wrote: "Of all of these fields, the learning of languages would be the most impressive, since it is the most human of these activities. This field, however, seems to depend rather too much on the sense organs and locomotion to be feasible" [3].

The ability to learn and use language is one of the most characteristically human traits. Many animals signal, but only people can express and understand an essentially unbounded range of messages. The technical term for the ability of human language to support all these messages from a few dozen alphabetic symbols is "compositionality."

At one level, compositionality is obvious and not at all controversial; to quote the online Stanford Encyclopedia of Philosophy: "Anything that deserves to be called a language must contain meaningful expressions built up from other meaningful expressions" [4]. This obvious property of human language can be described as *Manifest Compositionality*: The meaning of a complex expression is determined by the meaning *in context* of its constituents and how they are assembled. Although this is rarely made explicit, any notion of meaning and grammar presupposes a *language community* (LC); an LC entails significant shared culture as well as conventions of language.

While the compositionality of textual language is manifest, it has been extremely difficult to define formalisms for studying it in full generality. The quest for a formal characterization of compositionality has led to a variant, often called *strong compositionality*, which states that the meaning of a (possibly complex) expression is totally determined by its form and is *independent of context*. As Manfred Krifka says in the online MIT Encyclopedia of Cognitive Science, "In its strict version, this claim is clearly wrong" [5]. For example, even simple sentences like *John loves Virginia* are ambiguous. The meaning is very different if Virginia is a state, a woman, or his baby. Even the grammatical structure of an isolated sentence is ambiguous, as in "He saw the man with a telescope." The telescope could be in the man's hand. Also, in this example, both "He" (a pronoun) and "the man" (a definite description) refer to people who are not identified in the sentence.

Following the long-standing tradition of formal language studies, this review will focus on the meaning of public printed text, like this paper. Handwriting and speech add further dimensions, and personal interaction goes well beyond just language in communication. Again following tradition, I will concentrate on the conceptual level; connections to detailed neural models and experiments can be found in Feldman [6,7]. While individual language acquisition and language evolution are important related issues, they will be discussed only briefly in the final section.

Almost all current mathematical and formal research on compositionality is about the strong form, is explicitly framed in logic-based semantics, and makes no claim to dealing with the full range of meanings. The focus is on technical problems like quantifiers and scope issues, which have proved to be extremely challenging and productive. At a very basic level, strong compositionality is a desirable property for formal mathematics, including logic. The meaning of a mathematical expression (or computer code) should not depend on context, except in well specified

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

restricted ways. There is a large body of elegant and insightful work in this logic-based semantics tradition; I suggest Partee [8] as an entrée to these developments and Dowty [9] and Barker [10] for an overview of current efforts. Discourse Representation Theory (DRT) is a continuing effort to extend logic-based semantics and its computational descendents beyond the individual sentence [11]; DRT will be discussed in Section 6.

By way of contrast, expressions in natural language must be usable over the full range of human experience and therefore must be sensitive to context. There is also now overwhelming evidence supporting our intuition that, for people, meaning involves a lot more than truth conditions (Section 2). As one example among many, what is the truth conditional meaning of a recipe in a cookbook? In this article, I will review an extended, *embodied* theory of meaning and present a formal treatment of the (context sensitive) manifest compositionality of human language and thought.

The situation in science, where compositionality becomes the standard criterion of general laws, is somewhat intermediate. A formula like Newton's second law, $F = ma$, is quite general, but implicitly presupposes the context of an unchanging, freely moving mass with constant acceleration at speeds much less than that of light. In fact, much of the difficulty of learning formal science involves understanding which formulas are appropriate in a given problem context.

In natural language, the meaning of most expressions depends heavily on context, in contradiction to the strong compositionality position. Even the word *red* denotes different colors in phrases like: red hair, red face, red light, red wine, etc. This is a simple illustration of conceptual blending [12], which is one form of meaning composition. There are additional meanings of red in accounting, politics, etc., and local meanings like the name or nickname of a person or team. Sweetser [13] discusses a wide range of blending cases that do not fit strong compositionality. Also, as we saw above, many language expressions such as pronouns inherently refer to something in context that is not explicitly in the utterance. Standard formal treatments of compositional semantics declare all such matters beyond their scope. DRT [11] considers some, but by no means all, context effects.

There is now a considerable literature [14] documenting additional processing time and difficulty for English sentences that require the reader to infer meaning that is not explicitly mentioned in the utterance. For example, "The journalist began the article after lunch" is usually interpreted to mean that she began writing (not reading, editing, shredding, etc.), but this is not stated directly. It is not surprising that such utterances are harder, again challenging the strong compositionality position.

To illustrate a more fundamental limitation of strong compositionality, consider the example of the ditransitive construction in Mandarin Chinese. This follows this basic pattern: verb of transfer, a subject (the giver), and two objects (the recipient and the theme), similar to an English sentence like: "Mother gave baby rice." In Mandarin, even when addressed to children, any verbal argument can be omitted if it is available from context. English and other languages allow omission in special situations like the response to a question, but it is universal in Mandarin. In studies that we will discuss further in Section 5, Mok and Bryant [15,16] examined ditransitive utterances from the Tardiff Beijing Corpus [17,18] in the CHILDES database [19]. In only 6% of the cases were all of the arguments expressed; the agent was omitted 78%, the recipient 41%, and the theme 66% of the time. This pattern is widespread in Asian languages and is obviously another refutation of strong compositionality. The meaning of the sentence depends in part on concepts that are not expressed at all.

So, we are confronted with a dilemma – manifest compositionality is a touchstone of human language and thought and should be studied scientifically. But the current formal scientific treatments cover only (context independent) strict compositionality and cannot be extended to the general case. Our proposed embodied approach to formalizing the manifest compositionality of language has two components. First, we focus primarily on *meaning* and then apply the insights to explain compositionality of surface form.

Second, we require a *formal* treatment of embodied meaning that is consistent with scientific findings at all levels, from neuroscience through culture. Of course, no formalization can *solve* a fundamental scientific question, but we can provide a foundation for systematic and cumulative advances. Successful formalization in any domain depends on two pillars: conceptual understanding of the domain and adequate mathematical representation. I will suggest embodied cognition as the conceptual base and neural computation and Bayesian best-fit analysis as the appropriate mathematical notation for a comprehensive formal theory of the compositionality of language and thought.

All this is further complicated by the fact that different people have differing interpretations of most sentences and of many words (contested concepts). This is not normally stressed in formal treatments, but any notion of meaning and therefore compositionality must be with respect to a language community (LC). Even within an LC, individuals

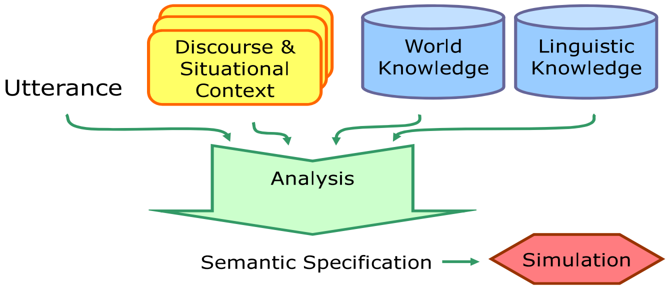


Fig. 1. Understanding an utterance in context.

inevitably have somewhat different interpretations of any utterance. Will an adequate theory need to encompass all human variation? Supposing, as we do, that meaning ultimately resides in a human brain, what can we say formally about written text – for example this article? It cannot be meaningless, because reading text does create broadly predictable effects. The key idea is that a grammar and the accompanying beliefs and desires describe the cultural conventions of communication and other conventionalized knowledge for an LC (Section 7).

The full conceptual system is different for each person. Despite this, people are able to communicate in some language because there is a shared *skeletal* belief system that is largely consistent within that LC. This skeleton is richly structured and complex, but is still a small subset of all the knowledge and experience in each person's mind. It is these shared skeletal beliefs that are communicated in language and whose compositionality needs to be formalized.

For example, there is a universal notion of possession that has many variants, but all of which share the schematic structure of roles for a possessor and a possessed entity. Possession is expressed in different languages by vocabulary items (e.g., own, my), by certain phrases (house of Jane), or by specific morphological conventions (Jane's house, case markings) – these are all *constructions* (Section 3). The individual and social *meaning* of ownership varies even more widely, but people within an LC normally understand when an ownership assertion or question is expressed. The Embodied Construction Grammar (ECG) formalism explicitly separates shared meaning that can be conveyed in a community from individual understandings which depend on personal beliefs, goals, etc.

Fig. 1 presents an overview of a contemporary model of embodied language understanding that will be used in this article. The top of Fig. 1 depicts the obvious facts that the interpretation of an utterance depends on linguistic and general knowledge as well as the discourse and situational context.

The bottom right of the figure reflects the deep insight that *mental simulation* plays a central role in understanding; we review the evidence for this below. The large arrow in the middle indicates a major hypothesis – language understanding can be divided into a general *analysis* phase which produces a Semantic Specification (cf. Fig. 5) that is common to an LC and independent of the particular beliefs and goals of the hearer. This is followed by an intrapersonal *simulation* phase, which updates the context and knowledge, among other things.

The ECG Semantic Specification formalizes the skeletal shared meaning of an LC. The *simulation* process, by contrast, always depends on properties of the hearer. This parcelization allows us to build a formal theory of shared language semantics and thus of compositionality.

As we discuss next, convergent results from several disciplines suggest a major role for embodied perceptual and motor experiences in language understanding. Language understanders automatically mentally imagine, or simulate, scenarios described by language. The mental simulations they perform can include motor detail at least to the level of the particular effector that would be used to perform the described actions, and perceptual information about the trajectory of motion (toward or away from the understander; up or down), as well as the shape and orientation of described objects and paths. The behavioral and neural imaging studies cited below suggest that these simulations involve brain mechanisms overlapping those responsible for perceiving the same percepts or performing the same actions.

Section 2 will review the now overwhelming evidence that meaning is indeed embodied and is inseparable from thought and experience. Section 3 describes how to formalize this shared knowledge, which arises from common genetics and experience. Section 4 illustrates how to formalize the compositionality of this conceptual knowledge, and Section 5 shows how this conceptual mechanism drives the manifest surface compositionality of language. In

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

Section 6, the formalization is extended to contextual matching and reference resolution. The final section talks briefly about language learning and change and a bit more about public and private meaning.

2. Embodied language

One of the most vibrant threads in contemporary Cognitive Science is the reformulation of cognition as inherently embodied or grounded. This encompasses all the relevant fields from neuroscience and computation to social cognition and philosophy. This article will focus on neural computation and language; current reviews from the perspective of experimental psychology can be found in [20] and [21].

Even within the field of embodied language, there are two largely separate areas of concentration. Cognitive linguists have, for several decades, focused on the fact that semantics (meaning) is not an abstract formal property, but is continuous with human perception, action, emotion, etc. We will review some of the behavioral and linguistic evidence for embodied meaning along with more recent findings from neural imaging.

A second formulation of embodied thought and language emphasizes the computational properties of the brain and how this constrains psychological theories. For the most part, these studies in neural computation (connectionist systems) have not dealt with language and abstract thought. For our purposes of explaining the manifest compositionality, both the cognitive and computational insights into embodiment are needed. A detailed account of this synthesis can be found in [6].

The notion that mental access to concepts is based on the internal representation of embodied experience is supported by recent brain research, which shows that motor and pre-motor cortex areas associated with specific body parts (e.g., the hand, leg, and mouth) become active in response to motor language referring to those body parts [22–24]. Using behavioral and neurophysiological methods, Pulvermüller et al. [25] and Hauk et al. [26] found that verbs associated with different effectors activate appropriate regions of motor cortex. In particular, Pulvermüller and colleagues had subjects perform a lexical decision task – they decided as quickly as possible whether a letter string was a word of their language, with verbs referring to actions involving the mouth (e.g., *chew*), leg (e.g., *kick*), or hand (e.g., *grab*). They found that the motor cortex areas responsible for mouth, leg, and hand motion exhibited more activation, respectively, when people were processing mouth, leg, and hand words. This result has been corroborated through Transcranial Magnetic Stimulation work [27]. Tettamanti et al. [28] have also shown through imaging that passive listening to sentences describing mouth versus leg versus hand motions activates corresponding parts of pre-motor cortex (as well as other areas).

Behavioral studies also offer convergent evidence for the automatic and unconscious activation of perceptual and motor systems during language processing. Work on spatial language [22,29,30] has shown that listening to sentences with visual semantic components can result in selective interference with visual processing. While processing sentences that encode upwards motion, like *The ant climbed*, subjects take longer to perform a visual categorization task in the upper part of their visual field (deciding whether a shape is a circle or a square). The converse is also true – downwards-motion sentences like *The ant fell* interferes with shape categorization in the lower half of the visual field. These results suggest that understanding spatial language evokes visual simulation that interferes with visual perception.

A second behavioral method [31] tests the extent to which motor representations are activated during language understanding. When subjects hear or read a sentence that describes someone performing a physical action, and are then asked to perform a physical action themselves, such as moving their hand away from or toward their body in response to a sentence, it takes them longer to perform the action if it is incompatible with the motor action described in the sentence. For example, if the sentence is *Andy gave you the pizza*, subjects take longer to push a button requiring them to move their hand away from their body than one requiring them to move their hand toward their body, and the reverse is true for sentences indicating motion away from the subject, like *You gave the pizza to Andy*. This interference between understanding language about action and performing a real action with our bodies suggests that, while processing language, we use neural structures dedicated to motor control.

A third method, used by Stanfield and Zwaan [32] and Zwaan et al. [33], investigates the nature of visual object representations during language understanding. Zwaan and colleagues have shown that the implied orientations of objects in sentences (like *The man hammered the nail into the floor* versus *The man hammered the nail into the wall*) affect how long it takes subjects to decide whether an image of an object (such as a nail) was mentioned in the sentence. When the image of an object is seen in the same orientation as it was implied to have in the sentence (e.g.,

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

when the nail was described as having been hammered into the floor and was depicted as pointing downwards), it takes subjects less time to perform the task than when it was in a different orientation (e.g., horizontal). The same result is found when subjects are just asked to name the object depicted. Zwaan and colleagues also found that when sentences imply that an object would have different shapes (e.g., an eagle in flight versus an eagle at rest), subjects once again responded more quickly to images of that object that were coherent with the sentence – images of those objects that have the same shape as they would have as described in the sentence.

In addition to all these imaging and behavioral results on the embodiment of individual word meanings, there is a growing literature on simulation semantics. From the theoretical perspective, formal logic and the related linguistic theories do not have a way of expressing or reasoning about detailed actions. Logical approaches have produced deep insights on truth-theoretical statements, but have not treated actions or sentences about actions, such as our first example, *Mother gave baby rice*. There are several well developed formalizations of action; we will employ Narayanan's X-schemas [34], which have been integrated into the ECG grammar used here. We will describe the formalization of X-schemas in Section 3 after presenting some of the evidence that people do mentally simulate actions when understanding sentences.

One powerful experimental method investigates whether sentences take longer to process when the scenes they describe take longer to mentally simulate. Matlock [35] demonstrates that the time subjects take to understand fictive motion sentences (sentences like: *The road runs through the* desert or *The fence climbs up to the house*) is influenced by how quickly one could move along the described paths. For example, a sentence like *The path followed the creek* is processed faster when it follows a paragraph describing an athletic young man who jogs along the path than when it follows one describing an old man who has difficulty walking all the way down the path. Similarly, characteristics of the path itself, like its distance or its difficulty to navigate, influence processing time in the same direction – the longer it would take the mover to travel the path, the longer it takes subjects to process the fictive motion sentence. This work once again suggests that processing language makes use of a dynamic process of mental simulation.

3. Formalizing embodied language

In order to present the outlines of a formal theory of manifest compositionality, we need to introduce a fair amount of ECG (Embodied Construction Grammar) notation. A more comprehensive introduction to ECG can be found in [36] and the full details are available through the NTL wiki – http://ecgweb.pbworks.com.

Embodied Construction Grammar is a formalism for specifying grammars that is being designed to simultaneously serve the following six functions:

a. A notation for the shared grammar and concepts of a language community.
b. A technical tool for linguistic analysis.
c. A computer specification for implementation of linguistic theories.
d. A front end system for applied language understanding tasks.
e. A representation for models and theories of language acquisition.
f. A high-level functional description for biological and behavioral experiments.

This review mainly exploits the first three functions; there is a brief discussion of acquisition in the final section. The relation between these higher level issues and detailed neural models and experiments is outlined in [6]. A recent articulating technical treatment of the neural modeling details is [7].

In ECG, construction grammars are specified using two basic primitives: **constructions** and **schemas**. Constructions are paired form constraints and meaning constraints. ECG is different from other construction grammar formalisms because the meaning constraints are defined in terms of embodied semantic schemas, such as those in Fig. 2, below.

3.1. Schemas

Schemas are used to represent a variety of conceptual structures, including image schemas, and frames [37]. Consistent with other analyses of such structures, schemas are defined as gestalt-like wholes with a limited number of internal parts, which are represented as **roles**. Crucially, rather than being defined as isolated, stand-alone structures,

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

```
schema ForceTransfer
   roles
     supplier
     recipient
     amount
```

```
schema Process
   roles
     x-net: X-net
     protagonist
```

```
schema EventDescriptor
   roles:
  profiledParticipant
  eventType: Process
  profiledProcess: Process
  profiledState
```

```
schema MotorControl
   subcase of Process
   roles
     actor: animate
     effector
     effort
     routine
  constraints
     actor ↔ protagonist
     routine ↔ x-net
```

```
schema ComplexProcess
   subcase of Process
   roles
   protatgonist  // inherited
     protagonist2
    x-net: @complexxnet
     process1: Process
     process2: Process
  constraints
     protagonist ↔ process1.protagonist
     protagonist2 ↔ process2.protagonist
```

```
schema ForceApplication
   subcase of MotorControl
   evokes ForceTransfer as ft
      roles
        actedUpon
            effort
      constraints
        actor ↔  ft.supplier
        actedUpon ↔  ft.recipient
            effort ↔  ft.amount
```

```
schema CauseEffectAction
  subcase of ComplexProcess
    roles
      process1: ForceApplication
      process2: Process // inherited
      causer
      affected
  constraints
      protagonist ↔ causer
      protagonist2 ↔  affected
      process1.ActedUpon ↔ affected
```

Fig. 2. ECG representation of several schemas.

these schemas are described as part of a larger lattice of schemas, with each schema having various types of specified relations to other schemas in the lattice. This reflects the complexity and interconnectivity of the conceptual network these schemas are being used to (partially) represent. In fact, any parcelization into separate schemas is just an approximation of the richly interconnected conceptual knowledge that is the substrate for meaning.

Various “primitive” schemas form a critical part of the lattice (Section 3). These hypothesized primitives reflect recurrent schematic commonalities in basic experiences. Such experiences are presumably shared by people, all of whom process them using some of the same basic functional networks in the brain. Therefore, these schemas are likely to be universally available to speakers of all languages, though they may of course be utilized in different ways by different languages. A fully defined grammar will also include schemas that represent recurrent commonalities in more culturally specific experiences. These schemas, akin to FrameNet frames [37], will also specify relations to other schemas in the lattice.

There are two additional classes of schemas that encode important functions in ECG: active schemas and compositional schemas. It has long been recognized that a major weakness of logical semantics is the inability to model actions. Since embodied language is crucially linked to action, the NTL group has developed and explored a fine-grained formal model of actions, called X-net [34]. This incorporates mechanisms for representing linguistic tense, aspect (e.g., *was walking* versus *walked*), repetition, interruption, and other detailed features of action. All this and more is needed to explain how people are able to perform detailed mental simulation of language input (cf. Fig. 1). The process schema in Fig. 2 shows that the ECG meaning of any process involves a protagonist and some X-net that models the details.

There are four ways to specify relations among ECG schemas: roles, sub-typing (through the **subcase of** keyword), evoking a structure (through the **evokes** keyword), and co-indexation and typing constraints. A **role** names a part of a

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

structure, and the **subcase of** keyword relates the construction/schema to its type lattice, allowing for structure sharing through (partial) inheritance; the MotorControl schema in Fig. 2 has a **subcase of** example.

Evoking a structure makes it locally available without imposing a part-of or subtype relation between the evoking structure and the evoked structure. For example, the concept shortstop only makes sense in reference to the sport of baseball, but it is not a subcase of baseball, nor is baseball a role of shortstop. The **evokes** operator models the neural fact that "shortstop" tends to activate the baseball frame. An example of evokes can be found in the ForceApplication schema in the bottom left of Fig. 2.

Like other unification-based formalisms such as HPSG (http://hpsg.stanford.edu/) and LFG (http://www.essex.ac.uk/linguistics/external/LFG/), ECG also supports constraints on roles. A type constraint (specified with a colon) constrains a role to only be filled by a certain type of filler. The double-headed arrow operator is used for co-indexing (binding) roles (↔). Both role constraints and bindings are depicted in the MotorControl schema of Fig. 2.

An important part of our story is that *compositional* schemas at the meaning level account for much of the form level compositionality in language. Several of the schemas that will be utilized later are described in this section, using Fig. 2 below.

At the top right of Fig. 2 is the EventDescriptor schema, which plays a central role in most ECG grammars. The meaning of any English clause is assumed to specify an event and a perspective from which it should be simulated – we will see several examples below. Fig. 2 also shows embodied process (Process) and force schemas (ForceTransfer, MotorControl, ForceApplication). The Process and ComplexProcess schemas represent structure common to a wide range of dynamic events such as motor-control actions, motion, and various object-related changes. The Process schema has two roles: a protagonist role for its single participant and an X-net role for the specific process associated with this participant. X-nets, originally inspired by research in biological motor control theory, represent detailed structure associated with dynamic actions and events [34,38].

ComplexProcess is a structure-building schema, defined independent of any specific processes, and specifies how two processes can be composed to form a single more complex process. It has two subprocess roles, called process1 and process2. Its X-net role is specified to be a "complexxnet" that integrates the X-nets of each of these subprocesses. The ComplexProcess schema shows how roles can be bound (co-indexed) – that is, required to have the same filler. The ComplexProcess's primary protagonist role (inherited from Process) is bound (using ↔) to the protagonist role of process1, and the secondary, protagonist2, role is bound to the protagonist of process2.

Compositional schemas comprise the structural foundation for building up meaning in ECG. The composition relation could be that process1 precedes process2, causes it, prevents it, etc. In our example of CauseEffectAction, process1 is an application of force that leads to an effect specified as process2. We will later describe in detail the CauseEffectAction example *He cut the bread* and then consider compositional extensions.

3.2. *Constructions*

So far, we have focused entirely on meaning and have said little about language form. This makes sense because we are proposing that the compositional nature of language is based on the underlying richness of thought. But we do need mechanisms to explicitly link form and meaning and these are named *constructions* in linguistics. Even the simplest grammatical constructions can be seen as linking a surface form to a conceptual combination. For example, the conceptual idea of possession can be expressed in English (but not, e.g., in Spanish) by a construction whose form is a nominal followed by 's as in "John's hand" as well as in other ways.

The Embodied Construction Grammar (ECG) formalism used in this paper is part of an ongoing tradition of research in construction grammar. Consistent with Cognitive Grammar and other construction grammars (e.g., [39–45]), ECG assumes that a grammar includes both lexical and non-lexical constructions. Furthermore, this set of constructions is structured both by (part-whole) constituency relations and by the relevant generalizations over form, meaning, and distribution. The goal is not to try to create the most "compact" grammar possible. In fact, we follow Langacker and Goldberg and posit usage-based grammars that might be better described as maximalist instead of minimalist (e.g., [46,47]). For one thing, the brain is much more constrained by processing time than by capacity to represent grammar. In addition, because the learning of the grammar is data-driven, there will be many small sub-regularities upon which generalizations are built [48].

Constructions are always pairings of form and meaning, and in ECG, this pairing is represented by a form block (defined by the **form** keyword) and a **meaning** block. Both the form block and meaning block of a construction can

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

ConstructionActive-Ditransitive-VP
subcase of Active-VP
constituents
v: Verb
obj1 : NP
obj2 : NP
form
v_f before $obj1_f$ obj1.f before obj2.f

meaning : TransferEvent
c.means ↔ v_m
$self_m$.recipient ↔ $obj1_m$
$self_m$.theme ↔ $obj2_m$
$self_m$.profiledParticipant ↔ $self_m$.giver

construction gei3
subcase of verb
form: Word
$self_f$.orth ↔ “gei3”
meaning: @give

schema TransferEvent
subcase of Event
roles
profiledParticipant
giver
recipient
theme

Fig. 3. A ditransitive construction example for Mandarin.

be typed. Form blocks can also have form constraints, and in simple lexical constructions, this is just the orthography. Fig. 3 has two example constructions as well as the TransferEvent schema which is used with them. We are using the *pinyin* transliteration of Mandarin and this is reflected in the construction for the word gei3. Here the form constraint specifies that the orthography of the gei3 construction is “gei3” by binding this string. The slot chain self_f.orth uses the **self** keyword to refer to the construction itself, and the subscript **f** to refer to the construction’s form pole. The Word schema has a role called orth to represent the orthography of a word. The **meaning** of this lexical construction is expressed as @ give, following the ECG notation for concepts that are in an external ontology (data base). The @ sign specifies the external ontology and, in this case, we assume that the deep meaning of gei3 is close enough to the English give, which is in the ontology.

The Active-Ditransitive-VP construction of Fig. 3 is more complex because it has three **constituents**. This shows how the ditransitive construction is represented in ECG, with ECG keywords in bold. The three constituents are a Verb and two NPs, represented by the local names v, obj1, and obj2 respectively. The **subcase** relation states that this Active-Ditransitive-VP is a subtype of Active-VP; subcase relations are important in preserving compositionality between constructions. Here, this enables a separate sentential or clausal construction to compose the Active-VP with a subject NP.

The ordering constraints among these constituents are expressed in the **form** block, which states that the constituent v must appear in the sentence before obj1, and obj1 before obj2. The meaning of this construction is a TransferEvent, whose **roles** are shown on the right in Fig. 3. The Active-Ditransitive-VP construction links its constituents to roles in this TransferEvent in the **meaning** block through identification constraints denoted by ↔. The recipient is identified with the meaning of the first noun phrase (obj1), and the theme is identified with the meaning of the second noun phrase (obj2). The crucial idea is that the meaning of an Active-Ditransitive-VP, which is a TransferEvent, *requires* all three arguments semantically, even though they may not all be present in the surface form.

3.3. Analysis and best fit

As an initial example of best-fit analysis using ECG, let us reconsider how Chinese children can understand a ditransitive utterance with missing arguments, like the Mandarin version of:

Example 1.

“You give auntie.”
“ni3 gei3 yi2”

People understand immediately that omitted argument was a theme (e.g., peach) in contrast to a sentence like “You give peach” in which the recipient is omitted – the question is how. The ECG best-fit analyzer embodies a theory of language understanding that can account for phenomena like this and many others.

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

This example also allows us to be more precise about the Semantic Specification featured at the bottom of Fig. 1. A semantic specification consists of all of the ECG schemas entailed by the input utterance plus the bindings of their roles. In this minimal example, the meaning can be approximated by a single TransferEvent schema; in general there will be a complex structure of linked schemas, as will be shown below in Fig. 5. Here, the full semantic specification requires bindings for the three roles: giver, recipient, and theme. In English, these are usually all explicitly mentioned and there are rules of grammar (constructions) that specify how the surface words bind to the schema meaning roles. The Analysis program [49,50], depicted as the large arrow in Fig. 1, uses Linguistic Knowledge to select the best-fitting schemas and bindings.

As we have seen, the situation in languages like Mandarin is more complex – one or more of the semantically required arguments may be absent from the surface form. The missing arguments must be filled in from Discourse and Situational Context plus World Knowledge, also depicted in Fig. 1. Discourse context refers to items mentioned in a previous utterance while situational context covers things that may not have been mentioned, but are part of the common ground. Matching an omitted argument inherently involves a best-fit process and, in fact, best-fit analysis is always required to deal with the multifold ambiguity of language.

We will return to this later, but it is already clear that best-fit will play a central role in any formalization of the manifest compositionality of language. The meaning of an utterance often depends on context – but not in an arbitrary way. An adequate theory of compositionality must include a treatment of *how* the analysis process merges contextual factors with the surface input to produce a semantic specification for an utterance. The particular best-fit analyzer program described here is consistent with several psycholinguistic findings, but is only an initial step toward a full theory (Section 6).

The ECG analyzer program [49,50] employs best fit to process sentences. The best-fit score of an analysis of a given utterance is a probabilistic metric which combines syntactic, contextual, and semantic factors. The syntactic component incorporates the combination of chosen constructions, their constituency relations, and the argument omission probabilities. The argument omission probabilities are approximated by the omission rates obtained from corpus data. The contextual component scores how well the referring expressions are resolved to items from context and how easily the omitted arguments are recovered from context. Finally, the semantic component scores the semantic bindings essentially by evaluating how well the frame roles are being filled [15,49].

To determine whether the omitted-theme analysis or the omitted-recipient analysis is more syntactically likely for the example sentence, we employ corpus statistics. It turns out that across the *gei3* sentences, the giver is omitted in 78.4% of the phrases, the recipient 41.2%, and the theme 66.0%. These statistics are used to estimate the likelihood of different omission patterns showing up in a sentence.

The omitted-theme analysis of Example 1 matches *gei3* (“give”) with the constituent v, *yi2* (“auntie”) with the constituent obj1 (i.e., recipient), and treats constituent obj2 (i.e., theme) as omitted. The likelihood of an omitted obj2 in the ditransitive construction above is 0.66. The omitted-recipient analysis, on the other hand, matches *gei3* as the v, treats obj1 (i.e., recipient) as omitted, and matches *yi2* with obj2 (i.e., theme). The likelihood of an omitted obj1 in the ditransitive construction is lower – 0.412. By the syntactic fit measure, then, the omitted-theme analysis has a higher score.

In Example 1, the speaker (the mother) is directing the child to give a piece of peach to auntie (the investigator who did the recording). The three of them are the only people in the room, but within the child’s surroundings are the peach that she may well be already holding in her hands and also a table. Both the omitted-theme analysis and the omitted-recipient analysis contain the referring expression *yi2* (“auntie”) which resolves well to the investigator in context. But the two analyses differ in how well the omitted argument is matched.

For the omitted-theme analysis, the peach is an obvious choice for the theme both because of its immediacy in the situation and of its fit as a theme. In this example, the context is strictly situational; there is no preceding discourse that could factor in. As a result, the omitted-theme analysis scores rather high in its contextual fit. The omitted-recipient analysis, however, is ambiguous as to whether the child, the mother, or the investigator should serve as the recipient. This ambiguity leads to a lower score for the omitted-recipient analysis.

The semantic component of the best-fit scoring reflects the semantic coherence of each analysis based on the likelihood of the frame role fillers. The two competing analyses of Example 1 both identify the addressee (the child) with the giver of the TransferEvent frame, but they vary with respect to the fillers for the theme and recipient roles.

The omitted-theme analysis picks auntie (the investigator) as the recipient and recovers the peach from context as the theme. By using resources such as FrameNet, our best-fit model is able to determine that a person is a good

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

recipient and a peach (or small physical object) is a good theme. In the omitted-recipient analysis, on the other hand, the investigator is constructionally chosen as the filler for the theme, and either the mother or the investigator is chosen from context as the recipient. Either of them is a good recipient, but a person is an unlikely theme. As a result, the omitted-recipient analysis is assigned a lower semantic fit score than the omitted-theme analysis.

The overall score of each analysis is obtained by combining all three component scores. Between the two competing analyses that we have considered here, the omitted-theme analysis is the obvious choice due to its higher scores in all three of the syntactic, contextual, and semantic components. Often, however, the choice among competing analyses is not so obvious. One analysis may have the best semantic fit, and yet a different analysis may have the best contextual fit. A feature of our probabilistic best-fit scoring mechanism is that it allows the analyzer to weigh these different factors in determining the best analysis. All of the examples in this paper use the probabilistic best-fit analyzer, and we will only discuss the best fitting analyses.

4. Conceptual composition

The central theme of this paper is that compositionality of language derives from an underlying compositionality of thought [22]. Section 2 outlined some of the evidence for embodied primitive concepts and Section 3 described how schematic versions of these are formalized in ECG. In this section, I discuss how complex and abstract concepts are formed as compositions of these primitives. The account is similar in form to the solution to a similar question in immunology. Animal immune systems are remarkably good at generating antibodies to combat novel antigens that invade the body. A raging question was whether this is a fixed process in which the killer antibody is selected from an innate repertoire or whether the system somehow manufactures a custom antibody, instructed by the intruder. The full answer is beyond the scope of this paper (and my knowledge) but the basic idea is clear. The immune system works because of a large number of primitive molecules that, in combination, can cover an astronomical number of possible antigens. The immunological primitives also evolve, but not fast enough to attack a new intruder. Gerald Edelman, who won the 1972 Nobel Prize for his research on the selection/instruction problem in immunology, has worked for decades to show how the same combinatorial principles can help explain the mind [51].

4.1. Conceptual primitives

There are many *direct* concepts involving body parts, actions, desires, experiences, etc. It is now clear (Section 2) that the neural representation of words and concepts concerning direct bodily experiences are based (at least in part) on the circuits that carry out the underlying action, emotion, perception, etc. This is not about some question of "innateness." We have known for decades that there is continual interplay between genetic and experiential (including cultural) factors starting from gestation and continuing throughout life [52]. There remain open scientific questions about exactly how these primitives (e.g., emotions) are encoded, but that is not the current concern.

However, there are also a large number of other, not obviously neural, potential conceptual primitives to explore and some fairly new experimental techniques that can help determine if a concept is primitive in our sense. Obviously enough, any concept that is learned is embodied somehow. The question is whether there is detectable neural encoding of mechanisms that help *organize* concepts and thus provide a basis for language and thought. Several developmentalists and linguists have suggested such possible conceptual primitives, but there does not seem to be any systematic collection of these suggestions. There should be.

Unsurprisingly, developmental psychologists have looked extensively at potential conceptual primitives [53–55]. Using strict standards of ontogenetic and phylogenetic continuity, Spelke and colleagues have identified four "core" cognitive primitives. To quote Spelke and Kinzler [56]: "These systems serve to represent:

> Inanimate objects and their mechanical interactions,
> Agents and their goal directed actions,
> Sets and their numerical relationships of ordering, addition and subtraction, and
> Places in the spatial layout and their geometric relationships."

The idea of "core" capacities has been extended to other domains such as emotion [57]. There are of course, many other concepts that are arguably primitive. Linguists including Jackendoff [58], Slobin [59], and Wierzbicka

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

[60] have suggested that words that appear in (almost) all languages suggest concepts that are likely to be universal and primitive. The most systematic effort along these lines is by Talmy [61,62], who outlines dozens of areas which have been grammaticalized in many languages and are therefore possible primitives. The following preliminary list of potential conceptual primitives draws on these and other sources.

Image Schemas

parameters of spatial cognition
action schemas – x-schemas, controller
goals, force-dynamics (causation)
parameters of parts & boundaries time

Social World

young/mature/old
authority, approval, help
value, exchange, obligation
theory of mind, perception and intention

Communication

speaker/hearer, direct/indirect
true/false
question, command, etc.

Grammaticalized Concepts

person, gender, age, agent, speaker
possession, mass/count, reflexives, instrument
Primal scenes/event types – transitive
Tense, aspect

Mental Operations

learning
matching, binding
mental spaces, mappings
simulation, displacement

General Logic

connectives, numbers
similarity, inference, uncertainty
part/whole, scales, magnitude
binding, variables, indefinites, generalization

The difference between primitive (universal) and cultural schemas is deep and important, but in NTL, this distinction plays no direct role in communication within an LC. However, it is crucial that we understand how new abstract and cultural concepts are developed by compositional operations using existing concepts.

4.2. Conceptual composition operations

There is extensive converging evidence for people's general ability to blend [12] and imagine (simulate) unbounded combinations of ideas. Before getting into the technical details, we can develop some intuition from examples like: "He walks like an angry ostrich." Even on a first encounter, we have no trouble imagining this scenario. Remarkably, you can readily substitute a vast range of actions, adjectives, and creatures in this example and mentally simulate each scene.

Formalizing this intuition, existing concepts can yield novel ones through compositional mechanisms like the following: conjunction (a zebra is a horse with stripes); modification (a llama is like a camel without a hump); abstraction (a vehicle is anything that can be used for transportation), and mapping (ideas are like objects), among others. These productive mechanisms can function through direct perceptual or motor experience (e.g., seeing an image of a zebra). But language can also indirectly ground conceptual learning. As shown above, language activates perceptual, motor, and affective simulation. This simulation itself constitutes experience that can form the basis for new concepts. The mental experience driven by language, and reproduced using the relevant neural circuits, is a sufficient basis for conceptual reorganization.

In fact, because of the brain's massive connectivity and spreading activation, concepts are never learned or activated in isolation – each of us boasts richly interrelated concepts. We are also continuously composing or "blending" concepts. We easily understand and imagine novel combinations like these. Fauconnier and Turner [12] are particularly interested in blends that combine different domains through mapping to a common space, like "trashcan basketball." They speculate that the human ability for complex conceptual integration was the key evolutionary advance that gave rise to language and thought.

Additional compositional mechanisms are required for describing complex actions. Among the most basic, and thoroughly studied, phenomena are linguistic *tense and aspect*. Consider the following four sentences:

(1) a. John is closing the drawer.
b. John is opening the drawer.
(2) a. John has closed the drawer.
b. John has opened the drawer.

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

Each of the four similar sentences entails a significantly different simulation and the differences would be similar (compositional) for many other statements of the same form. Recent experiments by Bergen and Wheeler [63] illustrate these differences as an example of the Action-sentence Compatibility Effect (ACE). Subjects were asked to read sentences like these and press one button (say $+$) if the sentence was meaningful and another (say $-$) if it was not. The crucial manipulation was whether the $+$ button was closer to or further from the body than the starting position. As predicted by simulation semantics, sentences like 1b that involve movement away from the body yielded significantly slower reaction times when the $+$ button was closer to the body. This is a well known embodiment result – it is the results for sentences like 2a, b that are crucial.

The only difference between sentences 1a and 2a is their aspect. In 1a, the closing process is described as ongoing (progressive), while in 2a it is specified as already completed (perfect). It follows from the simulation hypothesis that examples like 1a, b will induce mental simulation of the ongoing process and this, as we saw earlier, involves much of the same neural circuitry involved in actually moving the arm. If the required motion is in the opposite direction, interference should slow down the reaction. Bergen and Wheeler predicted that the ACE would be much less for examples like 2a, b where only the result is mentioned. They tested a variety of sentences on 70 subjects and the results were unequivocal. The progressive sentences yielded a strong ACE while the perfect sentences had no measurable effect.

Of course, our goal is not just to understand these phenomena, but also to formalize them and to eventually incorporate the results in language understanding systems. It turns out that linguistic aspect plays a central role in the pilot applications of ECG [64]. The key to formalizing semantic features of processes is an adequate model of action. For over a decade, the NTL (Neural Theory of Language) group has been working with an active process model [34,65] that incorporates simulation semantics and has been valuable in a wide range of theoretical and applied efforts. This model is incorporated in ECG as a specific kind of schema, called X-schema (X for executing) and the underlying realization called X-nets. These will be significant in the grammatical composition story of Section 5.

4.3. Metaphorical mappings

Another important mode of conceptual composition is through projection – describing a complex or abstract idea in terms of more basic embodied concepts. The best studied mechanisms for grounding abstract concepts is through mappings to them from concrete source domains. Abstract conceptual domains have long been known to be talked about in terms of concrete source domains, through linguistic metaphor [66]. For instance, English speakers (and speakers of many other related and unrelated languages) talk about ideas in terms of objects and knowledge in terms of object manipulation. Consider, for example, the phrases *I'm running out of ideas*, *I'm in the market for some new ideas*, and *I'm having trouble grasping the gist of the sermon*. Close analysis of texts reveals that for most abstract domains, non-expert language users exploit very little, if any, non-metaphorical language. The domain of ideas is a case in point. Ideas can be possessed, acquired, shared, chewed on, swallowed, recast, and worn out, among many other metaphorical construals. More generally, people always conceptualize new or abstract ideas in terms of ones that are already understood. Metaphor plays a large, if often unacknowledged, role in science [67,68].

A large body of research spanning the past thirty years provides convergent evidence that abstract conceptual domains are not only talked about in terms of concrete ones, but are actually thought about in terms of them as well. Early work in the Cognitive Linguistics framework [66,69] provides three main types of evidence that metaphor is not just describing-as, but conceptualizing-as. First, metaphorical language is systematic – when ideas are described as objects, considering the idea is manipulating the object; the considerer is always the manipulator, and the idea is always the object (and never the reverse). Second, this metaphorical language is productive. Concrete language is regularly used in novel, metaphorical ways, like the word *disintegrated* in *The new human stem cell research disintegrated under the light of scrutiny*. Third, not just language but also reasoning transfers from a concrete conceptual domain to an abstract one, through metaphor. So if *this theory is hard to get a grip on*, then we infer that this is due to a property of the theory itself – it's slippery or bulky – or to a property of the understander – they don't have sufficient mental skills to get their head around it. More recently, an important fourth type of evidence has appeared, behavioral evidence using tools from cognitive psychology, showing that language users activate concrete source domains when thinking about abstract target domains [70–73].

The case of conceptual metaphor shows not only how abstract concepts can be built up on the basis of concrete ones, but also how existing conceptual structures can be productively combined. The metaphorical grounding account

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

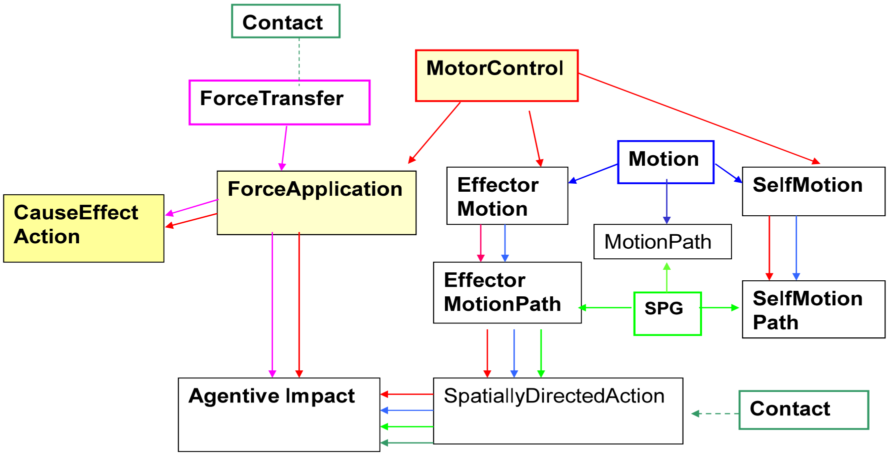


Fig. 4. A sub-lattice of conceptual schemas in the domain of force and motion.

sketched out above is insufficient to completely deal with some cases, like THEORIES ARE BUILDINGS (*Modularity is a foundation of the theory of generative grammar*; *These observations buttress the theory of natural selection*, etc.). There is no experiential correlation between the creation and structure of buildings on the one hand and the invention and organization of theories on the other. But Grady [74] has shown that the mappings by which theories are described and understood as buildings are *partial* – only certain aspects of buildings are mapped onto theories. These include the physical structure of buildings (foundation, support, and buttresses), and their persistent erectness, but not plumbing. The metaphor THEORIES ARE BUILDINGS is thus best seen as instantiating a combination of two primary metaphors – PERSISTENT FUNCTIONING IS REMAINING ERECT, and ABSTRACT ORGANIZATION IS PHYSICAL STRUCTURE. Each of these has a clear basis in experience. Many physical objects, like buildings, trees, chairs, and so on, function persistently only while erect. Many objects with complex physical structure also have associated functional organization – the legs are not only at the bottom of a table, but also serve the function of support. Put together through composition, these two primary metaphors produce a mapping whereby PERSISTENTLY FUNCTIONING ENTITIES WITH ABSTRACT ORGANIZATION ARE ERECT OBJECTS WITH PHYSICAL STRUCTURE. Metaphor and other maps have also been formalized in ECG [75], but this is beyond the scope of the current article.

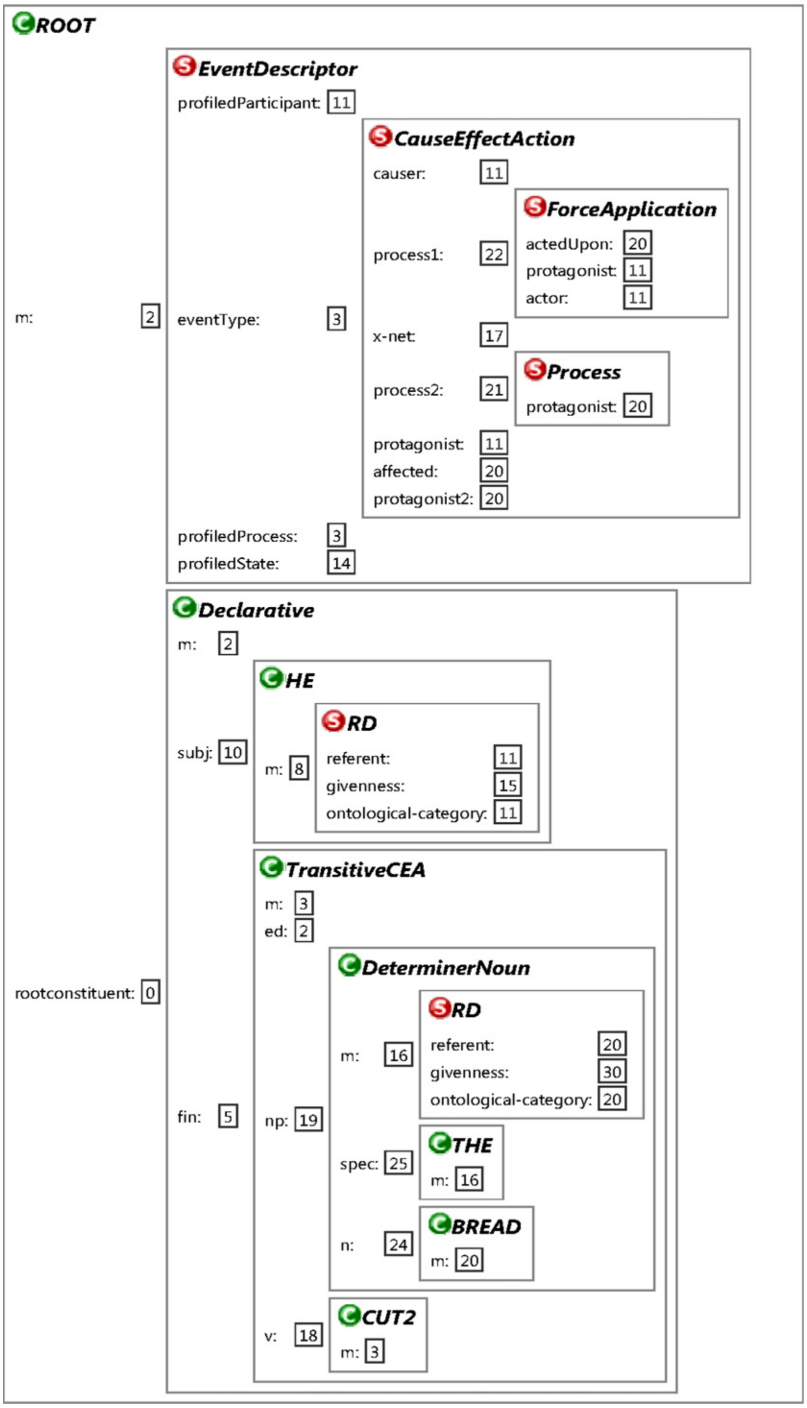
The conceptual structure of sub-domains can be quite intricate. Fig. 4 depicts a fragment of the ECG schemas used in Dodge's dissertation [76] and by Dodge and Bryant [77]. We will not describe all the relations in Fig. 4, but some are important for further examples. Notice first that MotorControl near the top is a parent of both Locomotion and ForceApplication, which is also a subcase of ForceTransfer. The dotted connection to Contact indicates that this is evoked by ForceTransfer. Fig. 5 presents the complete best-fit analysis of a CauseEffectAction for the example: *He cut the bread*. The schemas on the right of Fig. 4 are used to model the semantics of various kinds of motion, including self-motion, and some of these will be discussed below.

5. Grammatical composition

We are now (finally) able to show how the ECG formal theory of grammar explains the manifest compositionality of human language. Section 2 surveyed the evidence for embodied language and Section 4 above illustrated conceptual composition, which is the foundation of language. Section 3 described how ECG constructions and the best-fit analyzer map surface form to the conceptual level. The formal analysis of full compositionality involves the following three processes.

A) Analysis – This best-fit process is largely independent of context and individuals and is determined by the ECG grammar and ontology. It produces a SemSpec (Fig. 1).
B) Context Fitting – This happens in parallel with both analysis and simulation and contributes to both.
C) Simulation – Given a SemSpec and resolved references, this depends entirely on the beliefs and goals of the hearer.

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006



ROOT
m: 2
EventDescriptor
profiledParticipant: 11
eventType: 3
CauseEffectAction
causer: 11
process1: 22
ForceApplication
actedUpon: 20
protagonist: 11
actor: 11
x-net: 17
process2: 21
Process
protagonist: 20
protagonist: 11
affected: 20
protagonist2: 20
profiledProcess: 3
profiledState: 14
rootconstituent: 0
Declarative
m: 2
subj: 10
HE
m: 8
RD
referent: 11
givenness: 15
ontological-category: 11
fin: 5
TransitiveCEA
m: 3
ed: 2
np: 19
DeterminerNoun
m: 16
RD
referent: 20
givenness: 30
ontological-category: 20
spec: 25
THE
m: 16
n: 24
BREAD
m: 20
v: 18
CUT2
m: 3

Fig. 5. The ECG Semantic Specification (SemSpec) for “He cut the bread.”

We will focus first on process A – how the ECG best-fit analyzer produces semantic specifications for input sentences. Context fitting will be discussed in Section 6 and Simulation was described in Sections 1–3. The ECG analysis provides the clearest demonstration of how conceptual compositionality drives the manifest compositionality of surface form in language. Fig. 5 above is a screen image of the complete best-fit analysis of the simple sentence: *He cut the bread*. We will use this as a base for discussing how alternative compositional forms like questions, impera-

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

tives, etc., arise naturally. The example shown is analyzed against the partial grammar EJ1 [7], which specifies the form-meaning pairs for the four words in the sentence as well as the other constructions shown in the figure.

Before getting into the details, it is worth noting the significance of this result. Fig. 5 is a screen shot of the Bryant best-fit analyzer [49] in operation. This entails the formalization of the ECG schemas and constructions described earlier, a complex optimization program, and a specific partial English grammar (EJ1) that exhibits the kind of manifest compositionality that has not been previously formalized. The purpose of this section is to illustrate how the composition of language form derives from deep semantic (conceptual) composition as described in the previous section. The analysis is from the dissertation [76] and more information can be found there.

Looking at Fig. 5 in detail, first notice that the items preceded by an **S** all name ECG schemas from Figs. 2 and 4. Recall that a SemSpec consists primarily of a collection of schemas with bindings between various roles. The bindings are established when the analyzer successfully matches a construction; the bindings are specified by the ↔ statements in the meaning section of the construction. For example, back in Fig. 3, the first constraint in the MotorControl schema:

actor ↔ protagonist

specifies a binding between the meaning of the actor role and the protagonist role of the parent Process schema. One could try to explicitly draw all of these ↔ bindings, but it would soon become unreadable. ECG follows the standard convention of using boxed numbers to denote roles that are bound together. For example in Fig. 5, the tag **[11]** (∼he) denotes the profiledParticipant, as well as the protagonist in several schemas, and the referent of the first ReferentDescriptor (RD) schema. An RD encapsulates properties of entities mentioned in discourse – here just its ontological category and its givenness (whether or not it is new information). It is similar to the DRS in Discourse Representation Theory [78]. The second RD, tag **[20]** (the bread), is marked as old information and is, of course, of a different ontological type. Notice that **[20]** is the *actedUpon* of the ForceApplication, but the *protagonist* of process2, because it is the thing undergoing change.

The overall structure of Fig. 5 is an EventDescriptor (cf. Fig. 2). The EventDescriptor (ED) schema in ECG performs a function for clauses that is parallel to the RD for noun phrases. As noted above, the profiledParticipant, **[11]**, here corresponds to "he." The eventType is a CauseEfffectAction, as described in Fig. 4. This is a subcase of complexProcess (cf. Fig. 2) involving a ForceApplication (on the bread) leading to a process of the bread being cut. The second large box is constructional (marked **C Declarative**) and specifies that the utterance is a statement as opposed to a question, etc. The overall grammatical structure is a prototypical English transitive clause with a subject-verb-object configuration.

We will now present a number of sentences related to the example of Fig. 5 and show they all arise compositionally from underlying conceptual distinctions. The profiledParticipant role provides a lot of leverage for compositionality in the grammar. It allows for a simple semantic distinction between active and passive sentences and makes it straightforward to implement linguistic control (described below). As a consequence, the subject constituent is just like any other constituent in the grammar, and has no special status apart from the fact that its meaning is often bound to the profiledParticipant role. For example, a construction like the imperative without a subject (*Cut the bread*!) is not problematic because the profiledParticipant is bound to the addressee.

More generally, we note that: (1) A sentence not only instantiates lexical constructions, but also phrasal and other types of constructions. All of these constructions are meaningful, with meaning represented using schemas and bindings. (2) When these constructions unify, their meanings compose in a way that is consistent with the constraints specified in each construction. (3) In ECG this composed meaning is a semantic specification for a simulation (SemSpec).

The same construction can be instantiated in many different sentences, and should therefore compose with a variety of other constructions. For each sentence, the instantiated constructions should unify to produce a SemSpec that is consistent with our intuitions about that sentence's meaning. Similarities and differences in sentence meaning should be reflected in their SemSpecs. In addition, the ECG lattice of constructions facilitates expressing generalizations across constructions.

The crucial point is that all of the variations in constructional *form* exist to capture nuances of *meaning* as represented in the SemSpec. Consider some further variants of the example in Fig. 5 and their resulting bindings. These examples illustrate how a linguist can use ECG to formalize complex compositional form-meaning relationships.

Using the partial grammar EJ1, the passive variant: *The bread was cut* (*by him*) is analyzed as a passive argument structure construction whose meaning is CauseEffectAction, and which specifies that the meaning of its verb

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

constituent is one of ForcefulMotionAction. In these respects, this construction is the same as the TransitiveCEA argument structure construction that was instantiated above for *He cut the bread*. However, the passive argument structure construction specifies that the profiledParticipant role is bound to the affected role of the CauseEffectAction, not the causer role. This is a prototypical example of compositionality, where a particular semantic distinction (perspective) gives rise to two related surface forms. Both the passive and the active argument structure constructions specify the same type of event, and both have a verb constituent that elaborates the causal action within this event. However, they differ on which simulation perspective they specify, with active specifying the perspective of the causer, and passive that of the affected item.

The bread was cut therefore instantiates a different (though semantically similar) argument structure construction from *He cut the bread*. Most of the other instantiated constructions are the same for both examples, including Declarative, a CutPastTense verb construction, and an NP construction for “the bread.”

Which bread did he cut also instantiates the same verb and argument structure constructions as *He cut the bread*, as well as similar nominal constructions. When the instantiated constructions unify, the resultant SemSpec is very similar to that of Fig. 5. Both specify the same eventType and profiledProcess, and in both profiledParticipants is co-indexed with causer. There are three differences. First, labels on the large discourse box are different (“Wh-question” versus “Declarative”). Second, as in the passive, the analysis specifies that the topic role is bound to the affected role of the CauseEffectAction, not the causer role. Finally, the RD for the bread is marked as undetermined, rather than old information.

A more complex case arises in sentences like: *He wants to cut the bread*, which are called control relations in linguistics. The strategy for analyzing control relations in the partial grammar EJ1 using ECG also relies on leveraging the power of the profiledParticipant role. First, specify a set of control verbs, such as *want*, whose meaning can be defined as involving a second “event” role. The meaning of *want*, for example, would be represented as a “wanting” schema in which a wanter desires that some type of event take place.

Next, specify an argument structure construction whose verb constituent is a control verb, and which has another constituent that is itself an argument structure construction. Thus, there will be two EventDescriptors in the SemSpec, one associated with the control argument structure construction itself, and the other associated with the constituent argument structure construction. For the current example, this second argument structure construction would be TransitiveCEA, the same argument structure construction instantiated in *He cut the bread*. In the resulting SemSpec, the control EventDescriptor describes the event of wanting something, and the second ED describes the thing that is wanted (in this case the cutting event) or is for the thing that is wanted. In addition, the profiledParticipant role of the main EventDescriptor is bound to the control verb’s protagonist role, which in the current example is the “wanter” role.

For the construction instantiated in this example, the profiledParticipant is also bound to the profiledParticipant of the constituent argument structure construction. Therefore, in the SemSpec for *He wanted to cut the bread*, the profiledParticipants of each event descriptor are co-indexed, and are co-indexed with the wanter of the wanting event, and the causer of the “cutting the bread” event. And, because *He wanted to cut the bread* also instantiates Declarative, the profiledParticipant is also be bound to the meaning of the subj constituent, “he.”

Many other examples of constructional compositionality are discussed in Dodge [76]. So far, all of the cases examined have been treated well by the methods described above. First, analyze the conceptual domain (here process, force, and motion) and design an ECG lattice of schemas that capture its structure (Fig. 4). Also, specify the linguistic schemas needed; here, this includes the Referent Descriptor (RD) and Event Descriptor (ED). Then work out the lattice of ECG constructions that cover the phenomena of interest, always starting from the meaning (SemSpec) of the examples. Finally, use the best-fit analyzer to test that everything works compositionally. Of course, this all requires deep linguistic understanding, but we are now in the promised position of having a formal treatment of manifest compositionality.

It is not important here, but cognitive linguistics in general and ECG in particular incorporate two additional primitives beyond schemas and constructions. One of these primitives, *Situation*, models mental spaces [79] and related phenomena. All of the examples shown here have a single SemSpec, but discourse often involves other times, places, minds, etc., and therefore multiple SemSpecs. The other primitive, *Map*, formalizes the relation between situations, like relating an actor to a role that he plays. Maps are also used to capture metaphors, like those discussed in Section 5. The existing analyzer can process restricted examples of situations and maps and this is described in [75].

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

6. Context and reference resolution

To this point, the article has not seriously addressed the issue of formalizing *context*. Recall that compositionality of language is manifest if the definition explicitly includes the discourse and situational context of an utterance and also that strict context-free compositionality is demonstrably false. Therefore, any adequate formalization of compositionality entails formalizing context-dependence in language understanding.

Within the tradition of truth-theoretic semantics, the problem of context dependence has mainly been addressed by formal mathematical mechanisms. Two recent reviews summarize this approach.

> "Any model of interpretation can be made compositional if we are sufficiently relaxed about the nature of lexical meanings and/or the syntactic structures over which the compositional theory is defined" [78].

> "If you have been assuming meanings to be a kind of thing (call this the a-type) but discover that (strict) compositional interpretation needs access to some 'external' information b, that a does not supply, you can redefine your notion of 'meanings' to be functions of b-type things to a-type things" [9].

That is, one can include any required contextual information as the parameter of a function that will then yield the context-dependent a-type meaning as its value. This is called "dynamic semantics" and is essentially the same device used in trying to build a denotational semantics for functional programming languages; in that case one defines every function to operate on the giant vector of all variables. As you would expect, such technical tricks have no relevance to human thought and language.

The ECG parcelization into analysis and simulation, depicted in Fig. 1, provides a significantly different approach to context dependence for textual language. Since the semantic specification (SemSpec) is the meaning of an utterance shared by a language community, we can assume that the *situational* context is also shared. A language community is defined as sharing knowledge (including grammar) and the SemSpec is restricted to structures involving only shared knowledge. Section 3 described how to formalize this shared knowledge, which arises from common genetics and experience. Section 4 illustrated how to formalize the compositionality of this conceptual knowledge, and Section 5 showed how this conceptual mechanism drives the manifest surface compositionality of language. We will not focus here on examples where there is also a shared physical environment under discussion [80], but the extension is straightforward [49].

However, we have not yet seriously addressed the issue of *discourse* context, the dependence of meaning on previous (and sometimes subsequent) utterances. The standard examples are pronouns (e.g., she) or definite descriptions (e.g., the man). On any theory, some kind of *reference resolution* is needed and this depends on information not in the current utterance. For our purposes, it is convenient to characterize all issues of discourse context as selection and reference resolution. Selection, as in choosing the appropriate sense of a homophone like "bank", is already treated by the implemented ECG best-fit analyzer. In this section, I will outline the extensions necessary to model contextual reference resolution. Notice first that reference resolution and context processing in general must be part of the commonality across a language community. For successful communication the community must share not only background knowledge, but also similar methods for bringing this knowledge to bear on language, including reference resolution.

Reference resolution appears to be a distinct function complementing the analysis and simulation processes that are depicted in Fig. 1. In normal language processing, resolution happens rapidly and unconsciously. A large body of psycholinguistic experiments [81] suggests that potential antecedents are activated as soon as an anaphor is encountered. There is a wide range of factors that have been shown to influence which potential antecedents are activated and inhibited. Resolution, like all binding, is necessary for simulation but it often depends on context and meaning and so some inference is needed to find the links required for a complete semantic specification like Fig. 5. As we will see shortly, the inference involved in reference resolution is necessarily quite shallow – real time processing precludes any detailed simulation in the online resolution process. As with resolving alternative word meanings, there are occasions when automatic reference resolution processing fails and a conscious search is evoked.

Again, the basis for our formalization is the fact that communication relies on shared knowledge, which is packaged in ECG as a SemSpec. The crucial ECG element for reference resolution is the ReferentDescriptor (RD) schema. The example SemSpec in Fig. 5 contains two abbreviated examples of the RD, one for the pronoun "he" and one for

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

"bread." Let's focus first on the three RD roles presented in Fig. 5: referent, givenness, and ontological category. The latter is the most obvious: any referent descriptor will describe some category in the ontological lattice – this could be a person or food, but it could also be an action or event. The givenness role distinguishes between indefinite referents (e.g., a man) and definite ones (e.g., the bread) as well as finer distinctions such as between "this" and "that." Most crucially, the referent role in an RD links to the resolved *referent*. For example, in Fig. 5, the (resolved) referents might be a specific man mentioned earlier and a generic loaf of bread, assuming nothing else was known.

There is a linguistic research community actively studying reference resolution; its results are far more than can be reviewed here; Kehler [82] and Barrs [83] are a good introduction. The standard terminology links some text (the anaphor) to a prior entity (the antecedent). Although this does not appear in our examples, the RD mechanism can also handle references to a wide range of ontological types including actions, events, etc. For example, in "She likes swimming, but it tires her," the word "it" refers to an activity. Much of the standard literature ignores such event anaphoras and cannot extend to the general case. No existing treatment covers the full range of event anaphoras.

In the truth-theoretic semantics tradition, the strongest efforts on reference resolution have been developed with Discourse Resolution Theory (DRT). A good introduction to DRT can be found in [78] and a more computational version in [11]. The basic element of DRT, called a DRS, is similar to the ECG Referent Descriptor, but rather more complex than the simplified version used in this paper. In particular, a DRS can include several descriptors and logical relations among them. As the name suggests, DRT is concerned with resolving anaphors (e.g., pronouns) to appropriate antecedents in extended discourse. Within its self-imposed limitations, DRT research has yielded deep insights into several problems of context in language. In DRT, each new utterance must combine with the context, usually adding to relations to the context; this is a simplified form of the simulation semantics of ECG. In the current version of DRT [78], the reference resolution process is viewed as separate from the initial analysis and this is also consistent with the ECG approach. Finally, the DRT constraints on an antecedent are strictly logical and are called accessibility constraints.

DRT research has gone well beyond the simple cases and yielded proposals for dealing with quantifiers, including quantifiers over events. This, in turn, leads to a logical semantics for some issues of grammatical tense. Similarly, the ideas can be extended to handle some presupposition and mental attitude cases, but not the full range treatable by Fauconnier's Mental Space theory [79], which has been partially realized in ECG [75]. Many of the insights from this DRT work will be important in any formalization of context dependence, but there are also things that it cannot address. DRT shares the limitations of all truth-theoretic semantics, does not deal with situational context, and has only accessibility as an approximation to best-fit reference resolution.

Since detailed understanding in ECG is modeled by the simulation phase, it suffices to formalize how discourse context can be used to find the best-fit resolved referent for an RD in context. There is a somewhat delicate point that should be addressed here. While many (antecedent) resolved referents appear close to the matching RD, some can be arbitrarily far back in the discourse [81]. Of course, many referents in a discourse refer to communally shared knowledge that is not mentioned at all (e.g., Obama, Chicago). We propose to model distant antecedents as shared knowledge and assume that these are resolved by the more general ontological processes. There is no inherent conflict between the local discourse and more global-knowledge antecedent candidates; both can be evaluated by the best-fit analysis, as we will describe.

The key to modeling reference resolution in ECG is to exploit the probabilistic best-fit analysis process discussed in Section 3 in connection with the Mandarin version of "You give auntie." In fact, this is an example of reference resolution where there is no surface structure for one of the required semantic roles. The examples involving pronouns, definites, etc., discussed here are somewhat easier because the anaphor usually provides some information (category, number, gender) about the antecedent. In addition, the best-fit analyzer already includes the ability to compare alternative analyses for an overall best fit including syntactic, semantic, and contextual factors; this was also described in Section 3.

The current best-fit analyzer does include a primitive mechanism for reference resolution and this works well on a range of simple examples. The main limitation of the current system is that it has a simplistic model of discourse context and only considers a restricted set of possible anaphors and antecedents. There are much more sophisticated techniques in the literature that could be incorporated. The best-fit algorithm can evaluate alternative antecedents for overall compatibility using existing mechanisms. The current implementation uses forward prediction of the most likely constructions so it can handle at least some cases where the resolution comes after the antecedent. A much more sophisticated best-fit reference resolution mechanism is described in [84] and this could be incorporated into

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

the analyzer, although not without effort. The remaining issue is how the system (and people) can select the most plausible candidates to resolve a referent.

It might seem at first that there is an unbounded number of possible antecedents for an anaphoric reference such as a pronoun like “it.” There is a large literature on this, but we can make a good start from some general principles. Most basically, people understand language in real time and make analysis choices rapidly. In fact, subtle differences in reading time are one of the main diagnostic tools of psycholinguists.

Also, as you would expect, there is a strong bias toward recent items and constraints from the syntactic and semantic context of the anaphor. There are additional biases toward the subject or topic of the previous clause, etc. All of these can be accommodated in a constrained best-fit matcher [84], but real-time constraints make this program problematic as a psycholinguistic model.

All of this discussion has been about direct reference, finding an antecedent for a specific anaphor for the SemSpec. There is another important class of anaphora based on semantic frames such as those that are formalized in the FrameNet (FN) http://framenet.icsi.berkeley.edu/ project. We have previously used the baseball frame in some examples and it can help us illustrate frame anaphora. Often, the referent required in some contexts is heavily constrained by the frame. For example, in the baseball frame, “the pitcher” takes on a unique meaning and, in the context of a specific game, it resolves to a particular person. The FN theory explicitly includes representation for roles that are required for a given frame and which of these can be omitted in certain contexts. Again, there is a literature on frame-based anaphora [85].

In ECG, the reference resolution task in the analysis phase is formalized as finding resolved referents for all RDs, such as those depicted in Fig. 5, and evaluating these as part of the best-fit process. However, there are at least four distinct situations in which such resolution is not possible and some kind of proxy referent is required. Recall that reference resolution, like all online processing, is subject to severe time constraints and is sometimes incorrect. In the limited time available, some situations are best handled by inserting a proxy in the SemSpec and people are known to make the equivalent move.

These four situations are all familiar from ordinary language. For example, if someone says, “I saw your friend today,” the natural response is to *ask* which friend it was. More generally, one way to analyze an RD is to mark it as something that needs to be questioned. A second common situation arises in sentences like our earlier example, *He sliced the bread*. In many contexts, there will be no prior mention of the bread and the appropriate referent is the ontology item for a generic loaf of bread. A related situation is where the referent is the category itself, as in “She likes bread.”

There are also inherently future references (e.g., the next U.S. President) that must be treated in the SemSpec as currently unresolvable. Finally, since the analysis process is a preliminary step, it is feasible to allow a limited amount of unresolved reference in the SemSpec, assuming that the simulation phase can resolve it, possibly by trying the alternative simulations. For example, *He used it to slice the bread* does not specify the instrument, but we know that many potential antecedents are ruled out.

In addition to these four conditions, it will sometimes be true that a resolution problem is difficult enough to become conscious and possibly require significant problem solving; this is also occasionally true in other language understanding tasks such as word meaning.

The current theory also considers *metonymy* as an instance of the same kind of reference resolution problem. Metonymy is a standard mechanism of language where one item (usually a noun) is used as a way of referring to another [86]. There are many conventional metonymies such as using Washington to mean the U.S. government or, more generally, using a place to represent an institution based in that place. Fixed metonymy patterns like these can be directly represented in grammar, and the best-fit analyzer could score both the direct and metonymic referent and compare. But there are also rather more general metonymies like using a part to mean a whole (e.g., wheels for car) and these require an inference process similar to the basic frame anaphora discussed above.

In summary, the formalization of context effects is required for a formal treatment of the manifest compositionality of human language and thought. The crucial process is reference resolution – isolated sentences do not have unique meanings, but can be interpreted in context if all the ambiguities and references are resolved. Reference resolution involves two phases: selecting candidate antecedents and choosing among them. The best-fit analysis, which is required anyway, can evaluate alternative choices in context. Selecting potential antecedents is a hard, but not intractable, problem because there are multiple constraints and, in some cases, proxy solutions.

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

7. Conclusions and research issues

Compositionality is a defining characteristic of human language and thought, but has been difficult to formalize. For largely historical reasons, essentially all current research on formalizing meaning is based on truth-theoretic semantics and considers only a fraction of the problem. Significant progress is being made on a number of issues, but, in order to understand and formalize our manifest ability to understand novel utterances, both the definition of meaning and the technical machinery must be revised.

Section 1 outlines several of the important issues not currently within the purview of formal semantics. Section 2 surveys some of the evidence that, in people, meaning is embodied and is inseparable from perception, action, emotion, social cognition, etc. This suggests that we need a formalization of meaning that can capture this dependence.

Section 3 exploits decades of insights from cognitive linguistics to show how the required embodied semantics can be formalized at a schematic level, while respecting the fact that each person has individual interpretations. The second part of Section 3 extends this mechanism to formalize embodied *constructions*, which specify the relation between surface form and the schematic level Semantic Specification. The final part of Section 3 describes best-fit analysis, which is required to formalize the context dependent and probabilistic nature of human language.

Section 4 begins the exposition of compositionality in the ECG paradigm. It surveys the wide body of evidence demonstrating that conceptual compositionality is the core competence. We need to be able to freely form unbounded conceptual combinations and a great deal is known about how we do this. Given this conceptual ability, linguistic compositionality is relatively simple to explain, as Section 5 tries to show. The technical focus of Section 5 is the demonstration that the ECG construction lattice is a good mechanism for representing the compositionality of linguistically interesting cases. The best-fit analyzer provides tools for building and testing such grammars.

The paper to this point, as well as ECG and most other formal semantics research, has focused on isolated sentences. Human language obviously depends on both discourse and situational context and Section 6 surveys the state of the art on context. The best-fit mechanism is required for a formalization of context, but does not suffice. The main additional requirement has been extensively studied as "reference resolution." There has been some excellent work on various aspects of this problem [83], but the general case remains open.

This final section concludes with brief discussions of two important related issues: learning and the social aspects of language.

7.1. Learning and change

For simplicity, this article has completely backgrounded the issues of language learning and change. But, of course, all of the magic has to be developed anew in the mind of every person and, for some people, learning never stops. The NTL/ECG theory explored here has been developed with a significant focus on learning at all levels and appears to be a good formalization of the facts of embodied learning. Let's retrace the structure of the article and see how the learning story fits in.

Sections 1 and 2 outline the evidence for embodied language, much of which comes from developmental studies. Children learning about the world (and how to communicate about it) start first with concepts and words that are grounded in their direct emotional, perceptual, and motor experiences. First words vary significantly across individuals, but most English-speaking children's first words consist predominantly of concrete nouns, like *truck* and *ball*, and social-interactional words, like *up* and *more* [87,88]. It's obvious that concrete nouns are grounded in direct experience, but importantly, social-interactional words are equally bound to embodied experience. A child who utters *up!* is not soliloquizing on the existence of "upness" in the universe – he is using the word to label (often to bring about) a particular type of experience, where he is lifted. Often children also acquire early concrete verbs like *push* and *sit*. The NTL group has produced detailed simulation models of embodied word learning [6]. The formalization at the level of conceptual schemas (Section 3) plays a crucial role in these models. In addition to concepts that directly label their experience, children have pre-linguistic organizing schemas such as support, containment, and source-path-goal [53].

It's only once children are well along in their development of embodied words that they begin to develop language for abstract, distant, or general concepts [89]. Conceptual development progresses in the same way, with concrete and directly experienced concepts leading the way for greater complexity, as discussed in Section 4. There is extensive research on the development of conceptual composition, the most pertinent of which is the work on metaphor.

In the simplest cases, the two metaphorically linked domains are aligned in experience, and can thus become associated [66,74]. For instance, quantity is a relatively abstract domain, especially when applied to concepts like power, love, and social capital. But in early childhood experiences, as throughout life, quantity of physical entities varies systematically with concrete, perceptible correlates. In general, the more liquid in a container, the higher level of the liquid; the more objects in a pile, the higher the pile. The systematic correlation between a concrete, perceptible cue (physical height) and a more abstract and subjective one (quantity) leads the learner to scaffold the conceptual and linguistic structure on the concrete base. As the learner subsequently develops, the two domains are pulled apart – adults know that abstract quantity does not always correlate with physical height. In summary, concrete concepts are learned through schematization over direct experiences and abstract concepts are indirectly grounded through co-experience with concrete ones, or through compositional mechanisms that produce them on the basis of previously grounded ones.

The most powerful applications of the NTL/ECG theory to language development are in the domain of grammar learning. This is a topic of enormous activity and public controversy since it is used as a proxy for the nature/nurture debate [90]. All this is well beyond the scope of the current article, but is discussed in [6].

In this article, the formalization of constructions (Section 3) is the foundation for the ECG grammar learning models [91,48,16]. We have already discussed the Mok and Bryant work on using best-fit to analyze Mandarin utterances with omitted words. In fact, this was just a preliminary effort toward modeling how children learn Mandarin grammar when they rarely hear complete utterances. It is simpler to start with Nancy Chang's model of English grammar learning [92].

The basis for ECG acquisition models is the fact that a grammatical construction pairs linguistic form with embodied experience. Children always learn grammar in a context where they partially understand the situation. A child who already understands a scene conceptually and hears a sentence about it only needs to hypothesize what about the linguistic form that licenses the known conceptual composition. The model systems use best-fit analysis (Section 3) to match the constructions it already knows to the current situation. It tries to guess additional constructions that could map unexplained words to aspects of the scene that haven't been mentioned. Of course these initial guesses are often wrong, but the programs (and children) refine their grammars with experience. Interestingly, the grammar learning programs use a variant of the best-fit algorithm to refine the grammar choices. One can evaluate the fit of a grammar to a corpus of examples by summing the best fit scores on all the examples.

Mok [93] has extended this grammar learning to Mandarin Chinese, using the techniques described earlier in this article. The key additional idea is that children know the required conceptual frames before learning grammar and can thus match constructions with missing arguments, assuming the missing information is available in context, as we demonstrated. All of this stands as further support for the NTL proposals of embodied constructions, conceptual compositionality, and best-fit analysis. The grammatical compositionality of surface form described in Section 5 arises naturally from this learning procedure.

There is also a vast literature on language evolution and change that is beyond the scope of this article. Computational accounts of language evolution that deal with conceptual grounding and compositionality include [94–97].

7.2. *Public and private knowledge*

A standard criticism of embodied and neural theories of language is that they focus entirely on individuals and ignore the myriad social aspects of language. This is a valid concern and we end this article with an examination of the most critical question, the relationship between private and public knowledge.

If meaning is embodied, what is shared meaning – for example, the meaning of text like this article? This is a core question for any theory of language and our general suggestion is:

Shared meaning is partial and is a social construction of a language community based on similar genetics and experience.

This is independent of any particular theory, but is not vacuous since it eliminates any requirement for Platonic notions such truth, number, etc. The limited expressiveness of language is clearly articulated in [98]:

> "Language evokes ideas; it does not represent them. Linguistic expression is thus not a straightforward map of consciousness or thought. It is a highly selective and conventionally schematic map. At the heart of language is

> the tacit assumption that most of the message can be left unsaid, because of mutual understanding (and probably mutual impatience).”

Still at this general level, we know that everyone is a member of multiple language communities (LCs) and that LCs can be seen as a lattice with partial overlap. Everyone speaks differently in varying situations.

Much of the shared meaning of an LC is informal, but there are also many domains where LCs attempt to maintain formal coherence of thought and behavior. Among these are mathematics, science, law, baseball, traffic, social hierarchy, and ritual such as religious rites. Some bodies of shared meaning (belief systems), most notably science, have been remarkably powerful in enabling humans to affect the world in unprecedented ways. While all belief systems are socially constructed and evolving, they are not equally valid as ways of describing and predicting experience (*pace* some academics). Nevertheless, people communicate about all sorts of things and a theory of meaning should accommodate this richness.

NTL makes a number of specific assumptions about embodied individual meaning (Section 3) that have implications for shared meaning. One claim, common to all construction grammars, is that every element of language is a form-meaning pair. A second claim, coming from cognitive linguistics, is that the meaning side of such pairs can be expressed in terms of *schematic* structures that are skeletal versions of universal and cultural experience.

Language is inseparable from culture and cultural frames include ones like baseball, which is important in some LCs and totally absent from others. The difference between universal and cultural schemas is deep and important, but in NTL, this distinction plays no direct role in communication within an LC. This all suggests that *textual* shared meaning is communicated at the schematic level. Speech adds an additional dimension and personal interaction goes well beyond language in communication.

Another central tenant of NTL is *simulation* semantics – people understand language by imagining (or actually enacting) its meaning. This is the foundation of embodied meaning and is supported by a wide range of biological and behavioral evidence (Section 2). The remaining question is how people go from the schematic surface form of an utterance in context to this embodied meaning. The core idea here is that language understanding has two phases (Fig. 1). The first phase, *analysis*, converts an utterance in context to what is called the Semantic Specification (SemSpec), which is constant across an LC. Each individual then *enacts* the SemSpec in his own way.

To formalize these intuitions, we need an additional level of technical machinery: ECG and best-fit analysis as described in detail in Section 3. ECG is a precise formal notation for writing grammars that follow NTL principles. Crucial for our purposes is the claim that the best-fit analysis process and the resulting SemSpecs depend only on the LC and not on the individual understander. On the other hand, the selection and enactment of a SemSpec depends entirely on the particular beliefs, goals, etc., of the individual. The fact that we can share some, but not all, of the thoughts and emotions of others is called the problem of inter-subjectivity. In summary:

> *The NTL/ECG separation of language understanding into analysis and enactment phases provides a clean and well motivated distinction between the shared meaning of an LC and embodied individual meaning*.

This was all formulated from the recognition perspective, but the language production version is similar [49]. In anonymous textual communication, the best one can do is to attempt to induce a particular SemSpec in the audience, based on conventions of the LC that you share. If the LC fit is perfect, you should be able to generate text which will, with high probability, evoke the desired SemSpec, because your shared LC includes grammar, schemas, and best-fit rules. Even with a shared SemSpec, individual meaning will vary widely within an LC. Of course, there are many other issues in the social nature of language, but we believe that this formulation provides a foundation for linking individual and social phenomena.

Compositionality is a defining characteristic of human language and thought, but has been difficult to formalize. Formalization is necessary but not sufficient for progress in understanding human language and thought. Formalization brings all the advantages of the scientific method with unambiguous statements that are fit to the data and suggestions of further experiments. For language and thought, formalization extends the traditional goal of general (productive) insights. However, the fact that our minds are embodied places severe additional constraints on useful formalization. As I have attempted to show, any adequate cognitive theory must link with the supporting biology. More than this, human behavior is always context sensitive so there are strict limits on the predictive power of any formalization.

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006

From the neural perspective, knowledge/processing can be divided into neural connections, which NTL/ECG models fairly well, and the current state of activation, which is beyond any symbolic representation. As with all science, anything that we write down is only partial and approximate. No existing or proposed formalization can capture the richness of the massive parallel brain and the complexity of human culture, but we do need to take these seriously in theorizing.

Acknowledgements

Many useful suggestions were provided by the referees, Chris Switzer, and the indefatigable members of the NTL group.

References

[1] Semin G, Smith ER. Embodied grounding. Cambridge: Cambridge U Press; 2008.
[2] Jackendoff R. Foundations of language: brain, meaning, grammar, evolution. Oxford: Oxford U Press; 2002.
[3] Turing A. Intelligent machinery. In: Meltzer B, Michie D, editors. Machine intelligence, vol. 5. Edinburgh: Edinburgh U Press; 1948. p. 3–23. reprinted.
[4] Szabó ZG. Compositionality. In: Stanford encyclopedia of philosophy, http://plato.stanford.edu/entries/compositionality/, 2007. Last accessed: 5 May 2010.
[5] Krifka M. Compositionality. In: Wilson RA, Keil FC, editors. MIT encyclopedia of cognitive science. Cambridge, MA: MIT Press; 1999.
[6] Feldman J. From molecule to metaphor: a neural theory of language. Cambridge, MA: MIT Press; 2006.
[7] Feldman J. Ecological expected utility and the mythical neural code. Cogn Neurodyn 2009;4:25–35. doi:10.1007/s11571-009-9090-4.
[8] Partee BH. Compositionality in formal semantics. Oxford: Blackwell; 2004.
[9] Dowty D. Compositionality as an empirical problem. In: Barker C, Jacobson P, editors. Direct compositionality. New York: Oxford U Press; 2007. p. 23–102.
[10] Barker C, Jacobson P. Direct compositionality. New York: Oxford U Press; 2007.
[11] Hamm F, Kamp H, van Lambalgen M. There is no opposition between Formal and Cognitive Semantics. Theor Linguist 2006;32:1–40.
[12] Fauconnier G, Turner M. The way we think: conceptual blending and the mind's hidden complexities. New York: Basic Books; 2002.
[13] Sweetser E. Compositionality and blending. In: Janssen T, Redeker G, editors. Cognitive linguistics: foundations, scope, and methodology. New York: Mouton de Gruyter; 1999. p. 129–62.
[14] Hagoort P, Baggio G, Willems RM. Semantic unification. In: Gazzaniga M, editor. The new cognitive neurosciences. Fourth ed.. Cambridge, MA: MIT Press; 2009. p. 819–36.
[15] Mok E, Bryant J. A best-fit approach to productive omission of arguments. In: Proceedings of the 32nd annual meeting of the Berkeley linguistics society: theoretical approaches to argument structure. Berkeley, CA. Berkeley: Berkeley Linguistics Society; 2006. p. 35–55.
[16] Mok E. An embodied construction grammar approach to understanding and learning Mandarin Chinese. In: Boas HC, editor. Computational approaches to embodied construction grammar. Amsterdam: John Benjamins; 2010.
[17] Tardif T. Adult-to-child speech and language acquisition in Mandarin Chinese. Ph.D. dissertation, Yale University, New Haven, CT; 1993.
[18] Tardif T. Nouns are not always learned before verbs: evidence from Mandarin speakers' early vocabularies. Dev Psychol 1996;32:492–504.
[19] MacWhinney B. The CHILDES project: tools for analyzing talk. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
[20] Barsalou LW. Grounded cognition. Annu Rev Psych 2008;59:617–45.
[21] Glenberg A. Embodiment as a unifying perspective for psychology. Wiley Interdiscip Rev Cogn Sci 2010;1:586–96. doi:10.1002/wcs.55.
[22] Bergen B, Feldman J. Embodied concept learning. In: Calvo P, Gomila T, editors. Handbook of cognitive science: an embodied approach. San Diego: Elsevier; 2008. p. 313–32.
[23] Dodge E, Lakoff G. Image schemas: from linguistic analysis to neural grounding. In: Hampe B, editor. From perception to meaning: image schemas in cognitive linguistics. Berlin, New York: Mouton de Gruyter; 2005. p. 57–91.
[24] Gallese V, Lakoff G. The brain's concepts: the role of the sensory-motor system in reason and language. Cogn Neuropsychol 2005;22:455–79.
[25] Pulvermüller F, Haerle M, Hummel F. Walking or talking?: Behavioral and neurophysiological correlates of action verb processing. Brain Lang 2001;78:143–68.
[26] Hauk O, Johnsrude I, Pulvermüller F. Somatotopic representation of action words in human motor and premotor cortex. Neuron 2004;41:301–7.
[27] Buccino G, Riggio L, Melli G, Binkofski F, Gallese V, Rizzolatti G. Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. Cogn Brain Res 2005;24:355–63.
[28] Tettamanti M, Buccino G, Saccuman MC, Gallese V, Danna M, Scifo P, et al. Listening to action-related sentences activates fronto-parietal motor circuits. J Cogn Neurosci 2005;17:273–81.
[29] Richardson D, Spivey M, McRae K, Barsalou L. Spatial representations activated during real-time comprehension of verbs. Cogn Sci 2003;27:767–80.
[30] Bergen B, Lindsay S, Matlock T, Narayanan S. Spatial and linguistic aspects of visual imagery in sentence comprehension. Cogn Sci 2007;31:733–64.
[31] Glenberg A, Kaschak M. Grounding language in action. Psychon Bull Rev 2002;9:558–65.
[32] Stanfield R, Zwaan R. The effect of implied orientation derived from verbal context on picture recognition. Psychol Sci 2001;12:153–6.

[33] Zwaan R, Stanfield R, Yaxley R. Do language comprehenders routinely represent the shapes of objects?. Psychol Sci 2002;13:168–71.
[34] Narayanan S. KARMA: Knowledge-based action representations for metaphor and aspect. Ph.D. dissertation, UC Berkeley, Berkeley, CA; 1997.
[35] Matlock T. Fictive motion as cognitive simulation. Mem Cognit 2004;32:1389–400.
[36] Feldman J, Dodge E, Bryant J. Embodied construction grammar. In: Heine B, Narrog H, editors. The Oxford handbook of linguistic analysis. Oxford: Oxford U Press; 2009. p. 111–38.
[37] Fillmore CJ, Baker C. A frames approach to semantic analysis. In: Heine B, Narrog H, editors. The Oxford handbook of linguistic analysis. Oxford: Oxford U Press; 2009. p. 313–40.
[38] Feldman J, Narayanan S. Embodied meaning in a neural theory of language. Brain and Lang 2004;89:385–92.
[39] Langacker RW. Foundations of cognitive grammar, vol. 1. Stanford, CA: Stanford U Press; 1987.
[40] Fillmore CJ. The mechanisms of "construction grammar". In: Proceedings of the 14th annual meeting of the Berkeley linguistics society: grammaticalization. Berkeley, CA. Berkeley: Berkeley Linguistics Society; 1988. p. 35–55.
[41] Goldberg AE. Constructions: a construction grammar approach to argument structure. Chicago, London: U of Chicago Press; 1995.
[42] Kay P, Fillmore CJ. Grammatical constructions and linguistic generalizations: the what's X doing Y? construction. Language 1999;75:1–33.
[43] Croft W. Radical construction grammar: syntactic theory in typological perspective. Oxford: Oxford U Press; 2001.
[44] Langacker RW. Cognitive grammar: a basic introduction. New York: Oxford U Press; 2008.
[45] Michaelis LA. Sign-based construction grammar. In: Heine B, Narrog H, editors. The Oxford handbook of linguistic analysis. Oxford: Oxford U Press; 2009. p. 155–76.
[46] Langacker RW. Foundations of cognitive grammar, vol. 2. Stanford, CA: Stanford U Press; 1991.
[47] Goldberg AE. Constructions at work: the nature of generalizations in language. Oxford: Oxford U Press; 2006.
[48] Chang N. Learning embodied constructions. In: Boas H, editor. Computational approaches to embodied construction grammar. San Diego: John Benjamins; 2010.
[49] Bryant J. Constructional analysis. Unpublished Ph.D. dissertation, UC Berkeley, Berkeley, CA; 2008.
[50] Bryant J, Gilardi L. A cognitive model of sentence interpretation. In: Boas HC, editor. Computational approaches to embodied construction grammar. San Diego: John Benjamins; 2010.
[51] Edelman G. Neural darwinism. The theory of neuronal group selection. New York: Basic Books; 1987.
[52] Pinker S, Jackendoff R. What's special about the human language faculty?. Cognition 2005;95:201–36.
[53] Mandler JM. How to build a baby: II. Conceptual primitives. Psychol Rev 1992;99:587–604.
[54] Mandler JM. The foundations of mind: origins of conceptual thought. Oxford: Oxford U Press; 2004.
[55] Saxe R. Uniquely human social cognition. Curr Opin Neurobiol 2006;16:235–9.
[56] Spelke E, Kinzler K. Core knowledge. Dev Sci 2007;10:89–96.
[57] Barrett LF, Lindquist KA. The embodiment of emotion. In: Semin G, Smith ER, editors. Embodied grounding. New York: Cambridge U Press; 2008. p. 237–62.
[58] Jackendoff R. Consciousness and the computational mind. Cambridge, MA: Bradford/MIT Press; 1987.
[59] Slobin DI. The origins of grammaticizable notions: beyond the individual mind. In: Slobin DI, editor. The crosslinguistic study of language acquisition, vol. 5: expanding the contexts. Mahwah, NJ: Lawrence Erlbaum Associates; 1997. p. 265–323.
[60] Wierzbicka A. Universal semantic primitives as a basis for lexical semantics. Folia Linguistica 1995;29:149–69.
[61] Talmy L. Toward a cognitive semantics I: concept structuring systems. Cambridge, MA/London: MIT Press; 2000.
[62] Talmy L. The fundamental system of spatial schemas in language. In: Hamp B, editor. From perception to meaning: image schemas in cognitive linguistics. Amsterdam: Mouton de Gruyter; 2006. p. 199–234.
[63] Bergen B, Wheeler K. Grammatical aspect and mental simulation. Brain Lang 2009;112:150–8.
[64] Narayanan S. Moving right along: a computational model of metaphoric reasoning about events. In: Proceedings of the national conference on artificial intelligence (AAAI-99), Orlando, Florida; 1999. p. 121–8.
[65] Bailey D, Chang N, Feldman J, Narayanan S. Extending embodied lexical development. In: Proceedings of the twentieth annual meeting of the cognitive science society (COGSCI-98), Madison, Wisconsin; 1998.
[66] Lakoff G, Johnson M. Metaphors we live by. Chicago: U of Chicago Press; 1980.
[67] Jones R. Physics as metaphor. Minneapolis: U Minnesota Press; 1982.
[68] Brown TL. Making truth: METAPHOR IN SCIENCE. Champaign: U of Illinois Press; 2003.
[69] Lakoff G. The contemporary theory of metaphor. In: Ortony A, editor. Metaphor and thought. Second ed. Cambridge, MA: Cambridge U Press; 1993.
[70] Gibbs RW, Bogdanovich JM, Sykes JR, Barr DJ. Metaphor in idiom comprehension. J Mem Lang 1997;37:141–54.
[71] Boroditsky L. Metaphoric structuring: understanding time through spatial metaphors. Cognition 2000;75:1–28.
[72] Boroditsky L. Does language shape thought? English and Mandarin speakers' conceptions of time. Cogn Psychol 2001;43:1–22.
[73] Tseng M-Y. Exploring image schemas as a critical concept: toward a critical-cognitive linguistic account of image-schematic interactions. J Literary Semantics 2007;36:135–57.
[74] Grady JE. THEORIES ARE BUILDINGS revisited. Cognitive Linguistics 1997;8:267–90.
[75] Feldman J, Gilardi L. Extending ECG to communities, mental spaces and maps. In: Boas HC, editor. Computational approaches to embodied construction grammar. Amsterdam: John Benjamins; 2010.
[76] Dodge E. Conceptual and constructional composition. Ph.D. dissertation, UC Berkeley, Berkeley, CA; 2010.
[77] Dodge E, Bryant J. Computational cognitive linguistics: an embodied construction grammar analysis of transitive constructions. In: Boas HC, editor. Computational approaches to embodied construction grammar. Amsterdam: John Benjamins; 2010.
[78] Guerts B, Beaver DI. Discourse representation theory. In: Stanford encyclopedia of philosophy, http://plato.stanford.edu/entries/discourse-representation-theory/, 2007. Last accessed: 5 May 2010.

[79] Fauconnier G. Mental spaces: aspects of meaning construction in natural language. Cambridge, MA: MIT Press; 1985.
[80] Clark HH. Using language. Cambridge: Cambridge U Press; 1996.
[81] Nichol JL, Swinney DA. The psycholinguistics of anaphor. In: Barr A, editor. Anaphora: a reference guide. Oxford: Blackwell; 2003. p. 72–104.
[82] Kehler A. Coherence, reference, and the theory of grammar. Stanford, CA: CSLI Publications; 2002.
[83] Barrs A, editor. Anaphora: a reference guide. Oxford: Blackwell Publishing; 2003.
[84] Poon H, Domingos P. Joint unsupervised coreference resolution with Markov logic. In: Proceedings of the 2008 conference on empirical methods in natural language processing (EMNLP), Honolulu, Hawaii; 2008. p. 650–9.
[85] Poesio NM, Sturt P, Artstein R, Filik R. Underspecification and anaphora: theoretical issues and preliminary evidence. Discourse Process 2006;42:157–75.
[86] Markert K, Hahn U. Understanding metonymies in discourse. Artif Intell 2002;135:145–98.
[87] Bloom P. How children learn the meanings of words. Cambridge, MA: MIT Press; 2000.
[88] Tomasello M. First steps in a usage based theory of first language acquisition. Cognitive Linguistics 2000;11:116–35.
[89] Johnson CR. Constructional grounding: the role of interpretational overlap in lexical and constructional acquisition. Ph.D. dissertation, UC Berkeley, Berkeley, CA; 1999.
[90] Tomasello M. Constructing a language: a usage-based theory of language acquisition. Cambridge, MA/London: Harvard U Press; 2003.
[91] Chang N, Mok E. Putting context in constructions. In: Proceedings of the fourth international conference on construction grammar (ICCG4), Tokyo, Japan; 2006.
[92] Chang N. Constructing grammar: a computational model of the emergence of early constructions. Ph.D. dissertation, UC Berkeley, Berkeley, CA; 2009.
[93] Mok E. Contextual bootstrapping for grammar learning. Ph.D. dissertation, UC Berkeley, Berkeley, CA; 2009.
[94] Steels L. Grounding symbols through evolutionary language games. In: Cangelosi A, Parisi D, editors. Simulating the evolution of language. New York: Springer; 2001. p. 221–6.
[95] Perlovsky LI. Integrating language and cognition. IEEE Connections 2004;2(2):8–12.
[96] Brighton H, Smith K, Kirby S. Language as an evolutionary system. Phys Life Rev 2005;2:177–226.
[97] Fontanari JF, Perlovsky LI. Evolving compositionality in evolutionary language games. IEEE Trans Evolutionary Computations 2007;11:758–69.
[98] Slobin DI. Universal and particular in the acquisition of language. In: Wanner E, Gleitman LR, editors. Language acquisition: the state of the art. Cambridge: Cambridge U Press; 1982. p. 129–70.

Please cite this article in press as: Feldman J. Embodied language, best-fit analysis, and formal compositionality. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.06.006