Representing Inferences and their Lexicalization

David McDonald   DMCDONALD@SIFT.NET

Smart Information Flow Technologies, 14 Brantwood Road, Arlington, MA 02476 USA

James Pustejovsky   JAMESP@CS.BRANDEIS.EDU

Department of Computer Science, Brandeis University, 415 South Street, Waltham, MA 02453 USA

Abstract

We have recently begun a project to develop a more effective and efficient way to marshal inferences from background knowledge to facilitate deep natural language understanding. The meaning of a word is taken to be the entities, predications, presuppositions, and potential inferences that it adds to an ongoing situation. As words compose, the minimal model in the situation evolves to limit and direct inference. At this point we have developed our computational architecture and implemented it on real text. Our focus has been on proving the feasibility of our design.

1. Introduction

We begin this paper by laying out the problem we are addressing and our assumptions. Section 2 will descibe and motivate the new techniques and representations we are using. Section 3 provides an example of how they are used while Section 4 covers the same ground in much greater detail. We close with a section comparing our approach to similar ones and a summary of our claims.

1.1 Filling Gaps

One of the central facts about language is that speakers regularly omit information that their listeners fill in without any conscious effort. In other words, they leave *gaps*. Examples of such gaps are everywhere. Consider the following text:

> "...*a 14-year-old girl died in the Kurdish city of Sulaimaniya ... The rest of the family is in good health* ..."[1]

We effortlessly know that this is the family of the girl, even across the three intervening sentences in the full text. The writer could have said "*the girl's family*" but did not have to, knowing that readers would supply this information through inference.

Gaps like these are a pervasive and even essential component of language use: speakers appreciate what their listeners will infer from their knowledge of the world (e.g., children are presumed to have families) and from the communicative context that they share. This is one of the points of

---

1. This passage is excerpted from a January 18th, 2006, Aljazeera news article about the first bird flu victim in Iraq.

Grice's (1975) Maxim of Quality: do not be more informative than required . It is a central part of what makes a text cohesive.

A long-standing question, of course, is how this is done. How is our extensive body of background knowledge and inference organized? How do we deploy it so effortlessly? That is the subject of this paper, where we lay out some of our initial results from a recently initiated project into this question.

1.2 Speed Implies Structure

Psycholinguists have known for decades that language comprehension is immediate, incremental, and that it works on all levels at once: syntactic, semantic, discourse, and pragmatic (Marslen-Wilson, 1973). People interpret utterances word by word without noticeable delay. Recent work has shown that an event verb will activate its prototypical objects in just the time it takes to hear the verb and that this will influence the interpretation of later syntactic structures (Matsuki et al., 2011; McRae & Matsuki, 2009).

When cognitive psychologists explain this ability, they talk about people having *schemas* that organize their knowledge of ordinary things and events (Bartlett, 1932). This resonates with the ideas and mechanisms of frames and scripts that were developed in artificial intelligence and linguistics more than thirty years ago (Minsky, 1975; Schank & Abelson, 1977; Fillmore, 1976). These mechanisms encode knowledge about conventional types of events and situations that people know about or have experienced: birthday parties, presidental inaugurations, eating at a restaurant, etc. They provide expectations about what is likely to happen and what defaults to assume in order to account for things that must have happened but were not witnessed.

In areas of research such as neuroscience (Speer et al., 2009), or cognitive linguistics (Bergen, Chang, & Narayan, 2004), what a schema consists of or what it means, computationally, to 'activate' a schema and 'provide' expectations has different answers — it is usually not the point of their research. It is, however, the point of our own research. This paper describes our computational account of what schemas are, how they are activated, their mechanisms for controlling interpretation, and how they provide expectations, implicatures, and defaults.

1.3 The Importance of Knowledge

The knowledge-rich approaches of the 1970s and 1980s were abandoned by main-stream natural language research as part of the move to 'empirical' approaches that were made possible by the construction of large machine-readable text corpora and advances in machine learning (Church & Mercer, 1993). At about the same time, a shift to ever-larger projects increased the salience of the "knowledge acquisition problem" — that without a vast amount of knowledge, systems will be too brittle and will fail on anything outside of what has been expressly modeled. As a result, people working in natural language typically use shallow techniques that stop with just a description of what a text says and has none of the active, "fill in the gap" inferential capability that is critical for full, deep language understanding.

We agree that knowledge modeling is difficult. It is intellectually challenging to come up with conceptualizations that have the requisite sensitivity to context, the capacity for composition, and

associated expectations for actions and inference. But this background knowledge is absolutely needed if automated systems are to learn from reading or fully understand our instructions. We are not alone in this belief, as witnessed by the steady body of work by other groups, e.g., (Van Durme, Michalak, & Schubert, 2009; Montazeri & Hobbs, 2011). Moreover there are now substantial knowledge stores to draw on. In addition to Schubert's KNEXT, there are ConceptNet (Speer, Havasi, & Lieberman, 2008), FrameNet (Fillmore & Baker, 2001), and the long-term products of the CYC project (Guha & Lenat, 1993). Hence, we do not presume to do this by ourselves. Once our designs have been refined through testing on a realistic corpus against the series of prototypes we will implement, we intend to formalize our knowledge requirements and look for assistance from like-minded people in the language-centric part of the knowledge-representation community for follow-on collaborations.

1.4 What We Are Actually Doing

The focus of our work is on how inferences are marshaled from background knowledge when we use language. While it has old roots, our conception of how it is done is new. In order to focus our efforts, we have pushed to one side a set of issues that we know are important parts of any operational solution, but which now would just be a distraction.

- We are working from a corpus of written texts, not speech;
- We are not dealing with dialogue;
- We are not trying to acquire background knowledge automatically.

Instead we are working out how highly efficient, lexically triggered inference and expectation can happen at all. We are deliberately not yet invested in a particular ontology or a large knowledge store. We think it is more important to test and refine our computational machinery before drawing on the work listed above and working at a larger scale.

In the next section we lay out the elements of our architecture, and illustrate them in Section 3 with the example that has driven our design We follow this in Section 4 with a smaller but thoroughly implemented example that we walk through in detail. We conclude with a discussion of related work and our future plans.

2. Representation: Situations, Predicates, and Packets

Every cognitive architecture has a notion of *working memory*: some means of defining and delimiting what it will attend to and what it can be aware of at any given moment. Every architecture also has a *control structure*: a policy or mechanism that dictates what actions it will take and in what order.

In our architecture[2] – *C3* – our working context is a structured *situation*, in a sense close to that from situation semantics (Barwise & Perry, 1983; Devlin, 2006). We use a data-directed, event-driven control structure that adapts techniques used in our language analysis engine *Sparser* (McDonald, 1992; McDonald, 1996). We focus on the notion of a "situation type": a reoccurring pattern of events and participants. The situation semantics literature has instead focused on situ-

---

2. The name C3 stands for "the Compositional Construction of Context".

ations as a device that provide a denotation for a complex of events and participants — a static representation similar to Schubert and Hwang's notion of an episode in their episodic logic. The populated situation that accompanies an ongoing discourse supplies the information that is latent in the words of a text. In our view, situations hold the general world knowledge that perception unconsciously brings to mind. They supply the bulk of information that lies below the perceivable tip of the iceberg.

At its base, the situation holds representations of the entities, events, and predications that have been mentioned in the ongoing discourse. It provides a minimal model that consists of a set of typed structured objects. For example, if the text is "*a 14-year-old girl.*" then, when that phrase has been read, the situation contains representations of the girl, the age, and of the fact that the girl is described as being that age.

2.1 Lexicalized Pragmatics

In a lexicalized grammar, the terminals of the rules are specific words instead of lexical categories such as proper noun or transitive verb. We propose to lexicalize meaning and inference – to establish it directly from the incremental composition of the meaning of the words in a text without an intervening logical form.

The meaning of words, phrases, and meaning-bearing constructions is defined in terms of the set of entities, predicates, relations, propositions, or potential inferences they convey. Situations are created dynamically by composing these *packets* of content and inference as the words of a text are scanned. Most packets correspond to small individual categories or inferences, such as the affordances of a cup as a container or the consequences of a process being canceled. Packets are small because they are designed to compose with other packets to collectively define the suite of inferences that are active in a situation. Packets are activated singly or in groups according to what work they are designed to do and how and where they are triggered. The notion of *packet composition* is how we expect to satisfy one of the fundamental properties of language that have been recognized since the time of von Humboldt (1876): the ability to make infinite use of finite means.

2.2 Predicates Linked to Language

As a concrete example of a packet, consider the word *black*. It is the English realization of the individual in the ontology that is used to represent the color black (denoted as **black**), as opposed to other colors such as red or titanium white. Like all colors, it is associated with a two-place predicate that establishes a relationship between an entity that can have a color (tree leaves, cars, etc.) and the specific color **black**. We encode this predicates as

$$\lambda x_{hasSurface}[color\_of(x, black)]$$

where the type of object to which the predicate can apply is restricted: it must include the type **has-surface**. The object and the predicate together constitute the contents of the packet. When the parser scans *black*, these packets are introduced into the situation.

Every predicate in the ontology must specify the word or fixed phrase that expresses it and their linguistic properties.[3] The knowledge engineer adding colors to his conceptual model must indicate the word or phrase that names the color and that it has the syntactic patterns of a predicate adjective. For C3, we do this using the notation for simultaneously defining semantic categories and their realizations described by McDonald (1994).

2.3 Latent Predicates

When a phrase is fully instantiated, as in "*a black SUV*," the predicates receive values and establish predications. For example, the value of the color property of this SUV is bound to **black**. The meaning of substantive nouns or verbs will typically include a great many predicates, only a few of which will be present in a text and therefore explicitly represented as predications in the minimal model. The other predicates are *latent*. They may be relevant as the text continues; they may supply default assumptions that drive implicatures; or they may simply remain part of the background knowledge associated with the word, as we discuss in Section 4.3.

In C3 we treat predicates formally as a kind of *lambda variable*. These are structured objects that define a relationship between individuals of specific category, and are constrained in the range of values they can take, i.e., what the variable can be bound to (McDonald, 2000). This information is self-contained within the object that defines the variable – the category of individuals to which it applies, the restrictions on its possible values, and the default values that can be assumed in the absence of actual ones.

For example, if the participants of an event are physical objects then it is always the case that the event happened at some location, even if we do not know its identity. When the analysis of our initial example had only gotten this far: "*a 14-year-old girl died*," we knew that the death must have happened at some location, but we didn't know what that location was. The location could still be referenced, but only indirectly: "*where the girl died*" or "*the place where the girl died.*" Once the text continued, "*. . . in the Kurdish city of Sulaimaniya*," the latent variable that represented the location of the event was accessed and bound to that city. Note that this narrows the category of the location to **city**, and we would say "*the city where the girl died.*"

In our implementation, a composite category defines all the possible properties, relationships, and habitats (see below) that its instance individuals can have or can participate in, all implemented by lambda variables. When we introduce a packet into the situation, this potential becomes accessible, even when just a small part is present in the minimal situation model. It is accessible in that any later explicit reference to a predicate that is latent in an individual's composite category has already been linked to the individual. We employ a wrapper around all variables, effectively a programming trick, that lets C3 create an instance of each variable (potential predication) linking the relevant individual instantaneously in one step, at the moment the individual is introduced into the situation.

---

3. We use a Lexicalized Tree Adjoining Grammar for analysis and generation. A word's linguistic properties are established by indicating its TAG tree family or families (McDonald & Pustejovsky, 1985; McDonald, 1996).

2.4 Frames and Habitats

Packets are C3's building blocks. Most packets contain roughly the same amount of information as we intuitively associate with a single word (*black*, *cancel*). But of course there are relational structures that are considerably larger, structures that should be instantiated as a single unit but that have multiple parts and activities, such as an airport or a birthday party.

For C3, we represent these as *habitats* (Pustejovsky, 2013a). The notion of a habitat has its intellectual roots in two places. The first is that it is an extension and deepening of the well-established concept of *qualia theory* (Pustejovsky, 1995). We introduce a habitat into the situation all at once, but which aspect of it is in focus (which gets priority in dictating interpretations and making inferences) depends on what is in focus in the text being read, as we illustrate in Section 3.1. The term "habitat" deliberately plays on the ecological metaphor to guide intuition as to what should be included in a frame and what should not.

The other source for habitats is the knowledge representation techniques of classical AI: scripts for representing steriotypical events and episodic knowledge (Schank & Abelson, 1977), and especially the notion of a frame (Minsky, 1975).[4] Originally, frame theory emphasized the transformations that would occur as perspectives changed or scenarios progressed There was a focus on frame recognition and repair to account for variations. Action was tied to the creation of frames and to changes in their slot values via "attached procedures." Frames subsequently evolved into today's RDF triple-stores and weakly expressive description logics, in the process losing most of their value as a representation for background knowledge.

We have returned to something close to original conception of frames, but built from modern computational tools for abstraction and inheritance. Early knowledge-based comprehension research used pre-build monolithic frames; ours are assembled dynamically according to what is actually needed given the content of the text.

2.5 Indexical Functional Variables

The contents of a situation reside in a web of relationships and possibilities, most of them coming from the active habitats, others coming from the discourse relationships that structure the interpretation of the text, including relations that keep track of partial information as the text is being read. To represent this, we use a set of indexical functional variables similar to those in the Pengi system (Agre, 1988; Agre & Chapman, 1987). These variables designate constant, funtionally identical relationships within the processes of the system, while their values vary transparently to fit the moment-to-moment situation.

One of Agre's examples is the variable **the-cup-I-am-drinking-from**, which would be bound to whichever of the three cups that he kept in his office he was drinking from at the moment. The things he could do with this cup were always the same, while the identity of the cup would vary. The actions the system takes are stated once in terms of indexical variables – the presuppositions and significance of a functionally designated object is always the same. Actions are not dependent

---

4. A historical note. Minsky developed the concept of a frame during a seminar in the spring of 1972 that was dedicated to Newell and Simon's (1972) book *Human Problem Solving*. His starting point was Bartlett's (1932) notion of a schema.

on particular values, only on the function those values serve. Their actual values are managed automatically and transparently according to the situation at hand.

2.5.1 Pegs

In Pengi, the deictic variables were managed by its perceptual system. In our framework they are managed by the parser and identify the structure it has observed and the relationships it expects. In most instances an indexical such as **theme** or **new** will be bound to specific, typed individual, but since the situation is being updated incrementally as each word is scanned, there are always moments where a phrase is incomplete, its head and type not yet identified, but its impact on the situation still needs to be established. To do this we use Luperfoy's (1992) notion of a *peg*.

For example, at the point in the parse where we have read just "*a 14-year-old*," the indexical variable **current-np-referent** is bound to a peg that was created when the parser scanned the "*a*" and recognized that it was starting a noun phrase that would have a referent. The peg is a standin for the eventual referent and provides a place to accumulate predications. In this instance, we know that, whatever this referent may turn out to be, it is something for which it makes sense to have an age measured in years. The peg's properties are transferred to a regular individual once the head of the NP (*girl*) has been scanned. Section 4.2 provides another example of this process.

It is an interesting psycholinguistic question whether the earlier context has established the overall topic and narrowed the semantic field from which the referent of an incomplete phrase like "*14 year old*" will be drawn. The news article that this excerpt appeared in had "bird flu" in its title. Anyone familiar with the subject will know what types of individuals will be discussed and, given the age mentioned in the phrase, will presume that it will be a person. In other contexts, for example at a bar, the presumption might be that the 14 year old was a single malt scotch. Whether people use such pre-established semantic fields or wait a moment to hear the head word is an open question that could be tested in a well-designed experiment.

2.6 Representational principles and their consequences

We have arrived at a set of principles for the representation of world knowledge in C3. These are an overlay on an otherwise conventional system of categories and properties in a specialization lattice. The aim is to provide a flexibile link from language to the ontology while retaining the economy of only having to state axioms and relation types once. These principles include:

- Only add a category to the ontology if it makes a contribution, e.g., it adds predicates, state-change affordances, presuppositions, or defaults.
- No representation without realization; every category should correspond to some word, phrase, feature, or syntactic construction.
- Predicates are only defined once; they may be restricted to different values at different levels in the category lattice but they retain their identity.

In a conventional representation, there is a substantial distance in the specialization lattice between the particulars that appear in a text, such as a sport utility vehicle, which will be close to the bottom, and what we know about the vehicle, e.g. that it is a **container**, which is stated at a high level and applies to a great many things besides SUVs. It is difficult to use language in such a system. Our

need to have packets for domain-specific words that refer to general predicates and affordances (our *lexicalized pragmatics*) cannot be easily accommodated.

2.6.1 Unique variables

We chose instead to separate the realization facts (what words and construction are used) from the axiomatic facts (what predicates and operations apply and what follows from them). In C3 an SUV acts like a container because its category literally incorporates the **container** category and uses its variables to express the affordances available to its passengers and to state facts such as when one passenger gets out there is one fewer inside.

We do this by making all variables (predicates) unique. They are defined once, as one object in the representation, on a category as far up in the lattice as possible for maximal application. On more specific categories the variable will usually be restricted. For example the **contents** variable of **container** is defined there as a collection of an unknown number of entities of unknown types. When we move down to, say, **passenger-transporter** (see Section 3.2), the type of the collection is restricted to **person**. On a particular type of **passenger-transporter**, say **airline**, the restriction on the variable will be further restricted to incorporate the different roles of people on an airline.

The vocabulary is stated against these restrictions. Any packet that includes **container** adds to the situation model the fact that its contents are in one of two states, expressible as being *in* (*inside*) or *out* (*outside*) of the container, and have the affordance of being able to move between these states. But we say that we *take* or *pick out* jelly beans from a jar (they cannot move on their own). We watch a squirrel *climb out* of a garbage can (they can move on their own, and the movement involves ascending a height). When the variable is restricted to the category **person**, we refer to to them as *passengers* or by their role (*driver*, *pilot*, *stewards*), and they *go into* or *get out of* the container.

2.6.2 Pre-cached, "Composite" Categories

Allowing different local restrictions on the same predicate object lets us achieve an economy of expression for axioms, which is essential for working with large ontologies, while retaining flexibility in how to define packets of the vocabulary since realization facts can refer to restriction categories at very different levels in the ontology. But this comes at a cost, since any word with a rich meaning will have a packet that introduces dozens if not hundreds of latent variables (particularly for habitats) that will entail including a proportional number of categories.

We make this manageable by using what we call *composite categories*. We define them as a conjunction of regular categories. We then pre-cache the categories' variables (with their restrictions) to create a single computation object. The result has the behavior we would get by using ordinary inheritance, but with none of the costs of traversing the lattice to collect the variables and apply the restrictions.

While a a composite category often just collects the categories that are above it in the hierarchy, there is no requirement that it do so. Categories from very different parts of the ontology can be incorporated into a single composite. This makes for an ontology that is easier to maintain, since there is no requirement to force everything into a single lattice with coherent lines of inheritance. Com-
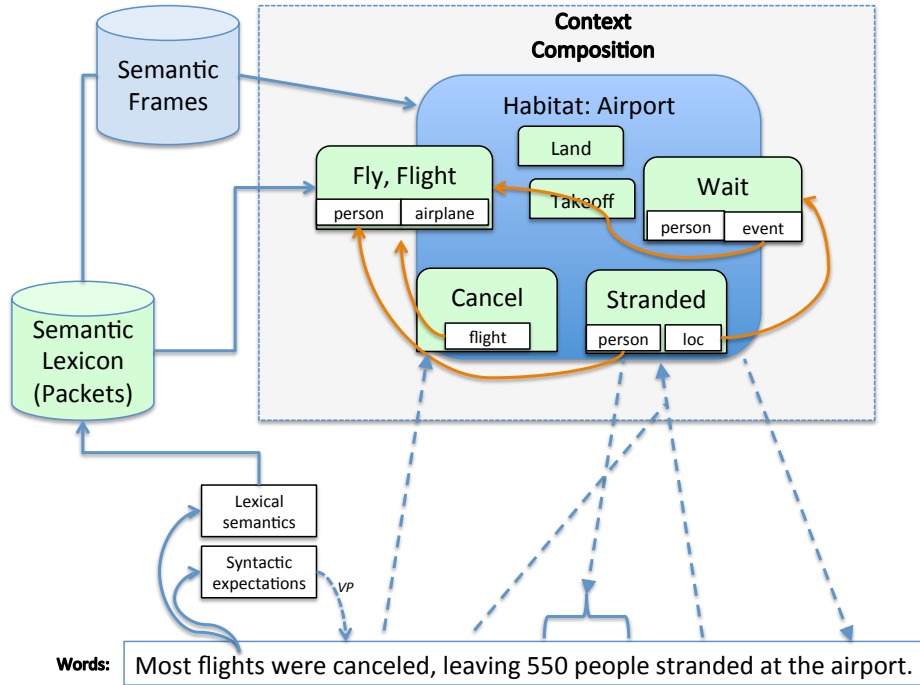
Figure 1. The C3 Architecture. As described in Section 3, the airport habitat includes a latent representation of its normal entities, roles, and activities. The C3 analysis incrementally brings some of these into focus, instantiating relationships and grounding otherwise anonymous text references such as the 550 people.

posite categories can be incorporated into other composites. When this happens, the incorporated composites are treated like macros that are unpacked inline and repackaged as a new class.[5]

2.7 The C3 Architecture

Figure 1 shows the basic framework of C3 using the example discussed in next section. Solid blue lines from the text trace the activation path up from the first part of the text to add packets (in green) or larger habitat frames (in blue) to the situation as a whole (outer box). Dotted lines show later additions to the situation (upward arrows) or inferred interpretations made by the situation (downward arrows). Orange arrows within the situation sketch relationships developed among the packets by binding variables.

C3's workflow begins with the perceived input; in our research this is the sequence of words in a text. Words are interpreted as they are reached by the parser and contribute packets of content of different sizes and function to a growing situation. This leads to the instantiation and assembly of highly structured sets of prototype relations and events, anticipated scenarios, and specific

---

5. We work in Lisp, and make heavy use of the multiple inheritance capabilities of the Common Lisp Object System (Gabriel, White, & Bobrow, 1991).

and prototypical individuals, places, and the like. The situation then governs the expectations and interpretations of words and phrases as the analysis continues.

The architecture proposed here assumes that utterances are interpreted incrementally, making use of inferential packets which drive the compositional construction of meaning. The result of the interpretation process is a minimal simulation of the situation denoted by the utterance. We turn now to an example of the interpretation procedure.

3. Creating and Applying a Situation

In this section we describe how the situation is established and drives inferences during C3's comprehension of an utterance. We focus on the following news article as our example text.

> "*Most flights from the Luis Munoz Marin Airport in San Juan to the Leeward Islands were canceled Monday, leaving about 550 people stranded at the airport.*" [6]

Taken just for its literal content, as with most of today's language understanding systems,[7] the result would leave many questions open. In particular, where did these 550 people come from, and why are they stranded? In the section below, we show how lexical semantic knowledge associated with the lemmas in this example direct our inferences towards "filling in the gaps" in the literal assertions from the text. We demonstrate how packets of information are formed from lexical items and how these compositionally build contextually salient inferences for this text.

3.1 Lexical Structures

Outside of a specific context, most high frequency words are ambiguous. Even once a word sense has been determined, there are still differences in logical perspective to sort out or metonymies to decode. We describe our approaches to these problems in this section.

Simple ambiguity. Consider the word *flights*, which has different meanings in different domains. It could refer to a *flight of stairs* or be part of a fixed phrase like *flight from stocks*. It could refer to a quantity of beer or champagne or it could be a nominalization of *flee*. A fully populated language understanding system would have all of those readings and more. In the context of this example it refers to an airline flight, but C3 must establish that fact before it can instantiate the **air-travel** habitat and activate its affordances.

We know from psycholinguistic studies that all of the senses of a polysemous word are available for about 250 msec after the word is read, and after about 500 msec, roughly when the next word has been read, only the contextually appropriate sense is available (Swinney & Hakes, 1976; Tanenhaus, Burgess, & Seidenberg, 1988). The context provided by the situation is sufficient for people to rapidly and unconciously disambiguate wordsthat are ambiguous in isolation like *flight*. The question is how does this happen. Various 'sublexical' techniques have been explored including marker passing (Charniak, 1983) and lateral inhibition (Cottrell & Small, 1983) though only in small systems.

---

6. This is a self-contained excerpt from a news article about the impact of Hurricane Earl on Puerto Rico (The New York Times, August 31, 2011).

7. Allen et al. (2007) and Clark, Harrison and Thompson (2003) report approaches that are notable exceptions to the generalization that today's parsers read just for literal content.

In C3, each kind of 'flight' that it knows about (for which is has a packet in its lexicon) has its own projection to the grammar, and will introduce its own semantically-labeled reading into the analysis when it is scanned, e.g., **airline-flight** and **fight-amount**. This mirrors the observed immediate activation of all the word's senses. When the next word is scanned, the word *from* in this example, it introduces its own projections, including its possibilities for composition in C3's lexicalized semantic grammar. This lets us use a simple disambiguation policy: only senses that can extend through composition with the phrases around them can have their meaning incorporated into the situation. The others are ignored. In this example we use **airline-flight** because the preposition *from* is part of the rule pattern that applies to 'flights' as movement (i.e. "flights *from* the Luis Munoz Marin Airport . . .").

Another possibility is that in an ongoing, established context such as news about a hurricane, the set of available readings for ambiguous words has already been narrowed to just those that are applicable in that semantic field. The psycholinguistic studies of lexical access (Small, Cottrell, & Tanenhaus, 1988) may well be based on stimulous conditions and probes that do not apply in the normal use of language between interlocutors who know they are in a shared situation. This would replace the problem of word sense disambiguation with the more realistic problem of recognizing the situation type. We intend to investigate this question in our future work.

Lexical entries in the generative lexicon. In Pustejovsky's (1995,2013b) Generative Lexicon theory , the lexical entry for a content word (as opposed to a grammatical function word such as *most* or *from*) encodes three kinds of information:

- Its **argument structure**, which spells out what arguments the word takes, how they are realized syntactically and govern semantic role selection;
- Its **event structure**, its class of event (state, process, transition) and how it structures its implicatures (Pustejovsky, 1991);
- Its **qualia structure**, the basis of logical polysemy, implicated in coercion and type shifting.

The argument structure is integrated into the rule sets of Sparser's grammar and helps with simple disambiguation. The event structure is part of the habitats that are added to the situation and provides a scaffolding for anchoring events and action sequences. The qualia structure organizes the applicable predicates and affordances.

Qualia and logical polysemy. The Qualia consist of four basic roles, each of which can be seen as answering a specific question about its associated object. Each contributes a complementary set of latent predicates to a word's meaning.

- **Formal roles** encode taxonomic information about the lexical item (the `is-a` relation). *What kind of thing is it; what is its nature?*
- **Constitutive roles**: encode information about the parts and constitution of an object (**part-of** or **made-of** relation). *What is it made of; what are its constituents?*
- **Telic roles** encode information on purpose and function (the **used-for** or **functions-as** relation). *What is it for; how does it function?*
- **Agentive roles** encode information about the origin of the object (the **created-by** relation). *How did it come into being; what brought it about?*

Most words have alternative readings that are characterized by different qualia: the newspaper you read (telic), the one you spill coffee on (constitutive), the one whose editorial opinions you disagree with (agentive). This distinction is referred to as *logical polysemy* (Pustejovsky & Boguraev, 1993). Once a content word has been narrowed to the domain where it has a specific meaning (simple disambiguation), the next step is to determine its qualia role, to disambiguate it logically.

The qualia role that applies in a particular instance cannot be determined independently of the rest of the context. If the text was *My flight just landed*, it would be the constitutive role, since we are talking about the airplane that the flight used and only physical things can land. If our flight was rescheduled, it would be the agentive role. All of these alternatives are part of the **air-travel** habitat — a frame that factors into different parts (incorporated habitats) according to which qualia is involved. In this instance of *flight*,[8] it is the telic role and it links to the portion of the habitat that organizes knowledge about flights as conveying people from place to place.

Metonymy. An important kind of inference is decoding what is actually meant when a general reference is used in place of a specific one. When "*the White House issued a statement*," it is not the building that did it but some spokesman and the identity of the spokesman is unimportant. This in happening in our example. The flights are "*from the Luis Munoz Marin Airport in San Juan to the Leeward islands . . .*"

The *from . . . to . . .* construction occurs in many situations, not just motion. In another context this could describe movement in stock prices. The fact that we are in an **air-travel** situation imposes a interpretation on the from–to arguments that they refer to airports. The airport in San Juan is given explicitly. The destination, however, is given as just a named location. An inference is required to convert from that to an airport. (Note that the writer could equally well have just said "*from San Juan*," in which case both references would require a metonymic interpretation.) C3 gets an airport from a place in several steps. Stated in conventional form, the rule would be something like this. Given the axiom (meaning postulate)

$$\Box[\forall f\text{:flight } \forall l\text{:loc}[destin(l, f) \rightarrow \exists a\text{:airport}[at(a, l) \ \wedge \ destination(a, f)]]],$$

we can devise a coercion operation such as

$$\lambda f\text{:flight } \lambda l\text{:loc}[destin(l, f)] \xRightarrow{coerce} \lambda f\text{:flight } \lambda l\text{:loc}\exists a\text{:airport}[destin(l, f) \wedge at(a, l)].$$

In Generative Lexicon terms, we are coercing the named location *the Leeward Islands* into the airport that is associated with that location. C3 can make that inference efficiently without needing search or matching because it uses the classical idea of *attached procedures*. Within the **air-travel** habitat, the origin and destination variables of a flight can only be bound to airports. This is already the case for the origin, but the destination is a place rather than an airport. Binding the destination to a place triggers a procedure to identify the needed airport in the situation. If it cannot be found then the rule simply creates a representation of 'the airports in the Leeward Islands' and adds it to the situation as the value of the destination.

---

8. Recall that the context is *"Most flights from the Luis Munoz Marin Airport in San Juan to the Leeward Islands were canceled Monday . . ."*

3.2 Habitats, Actions, and Composition

Airports have control towers, runways, taxiways, gates, and terminals. These are all available in the airport habitat. As we described in Section 2.4, these are entities and relationships that the habitat knows about, but they are latent rather than part of the situation's minimal model.

The principal activity at airports is air travel, and, if we ignore its personal aspects (making reservations, getting to/from the airport, buying food, shopping), the most salient aspect of air travel is the flights. Flights are also habitats. They have a plane (the equipment), a crew, passengers, baggage, and food. They are run by particular airlines, have a flight number, and travel from one airport to another.

In the telic reading of *flight*, the habitat includes a script that lays out the typical sequence of events and activities that constitute air travel. Airplanes are containers and they can move. Like any moving container, when they move (taxi, take off, fly, land), they convey their contents with them from their starting point to their destination. There are enough of these passenger-transporters in the world that they form a useful composite class: cars, buses, trains, bicycle-pulled carts, trucks, and others. This ensures that their common core is shared, particularly for our purposes the words that accrue to this level, such as *passenger*.

The interpretation of *flight* is as a process. There is a state of affairs that holds before this process starts and a different one after it ends. The principal difference between these two is in the location of the airplane and its contents: the passengers, their baggage, the crew. Before the flight leaves they are at the origin airport, afterwards they are at the destination airport. Any habitat like flight that involves scheduled process comes with the default assumption that once the process has started it will continue until it ends.

To represent the content of the first part of this text, we instantiate a flight habitat with values for the variables that we know. This adds to the situation a collection of an indefinite number of individual flights, where each of these otherwise unidentified flights originates in San Juan and terminates in an airport in the Leeward Islands. Each of these flights has a carrier and a flight number, a crew and a passenger manifest, but these are latent properties, just as we do not know the actual number of flights in the collection.

Canceling movement.   *Cancel* is an operator over processes: it modifies the situation rather than simply adding to it. Its syntactic configuration (as main verb) establishes that it applies to the value of functional variable **syntactic-subject**, i.e., the flights. Since the only qualia of flight that involves a process is its telic function of transporting its passengers from one place to another, that aspect of the **flight** habitat becomes central to the situation.

Applying the operator **cancel** to the flights cancels this process. To cancel a flight means that it does not start (the flight does not *take off*). This modifies the situation to reflect that fact that the conditions that held before the process would have started still obtain: the passengers who would have been on the flights are still at the San Juan airport, as are the crews and the planes.

Situation-driven binding.   In the last portion of the canceled flights example, we have a result clause

"*leaving about 550 people stranded at the airport*".

Given its form, the syntactic relation of this adjunct to its main clause tells us that this state of affairs (the stranding of the people) happened because of the event in the main clause (the cancelation of most of the flights). Being stranded is a habitat in itself, associated with air travel but not a part of it per se in the way as, say, losing one's luggage. The meaning of *stranded* is that there was an intention to move that has been blocked: The path of the passengers' expected futures has been interrupted. Note that the airport employees are not stranded, because they have a different role in the **air-travel** habitat, i.e., they work at the airport.

Inferences should be guided by what is salient in what is perceived – the text that C3 is interpreting and the situation model created for it. The cancelation brings into focus within the situation those elements that were most affected by it: the passengers, the air crews, and any other individuals whose intended future path of events was shifted. This salience makes it simple to interpret the two definite references in the result clause. Given the context provided by this situation, we can bind the referent of *the airport* to San Juan's Luis Munoz Marin airport because the **flight** habitat has already created properties for two airports (origin and destination). The origin airport is the more salient of the two because it is the one impacted by the cancelation.

Similarly, the *550 people* are resolved to be the only people who are made salient by the cancelation: the passengers and crew who would have been on the flights that did not take off — did not follow their intended, default future path.

This section has shown how simple it is to draw complex inferences in C3. We first recognize and instantiate the appropriate situation type ('activity at an airport'). That large habitat is focused on a particular qualia as the text is incrementally interpreted ("most flights"), and specialized through composition as C3 continues reading and introducing packets into the situation ("canceled"). This provides the context in which the identity of the '550 people' is immediately established, because they have the situation's role of `passengers` made salient by the cancelation of their flights. In the next section we will walk through this process in detail on a smaller, fully implemented example.

4. A Detailed Example

At SIFT we have access to a set of logs of actual text-chat collected from an Intelligence, Surveillance, and Reconnaissance team during the Empire Challenge 2010 military exercise. These are from a Ground-Based Operational Surveillance System team that was composed of three camera operators, an analyst, and a coordinator, all communicating over Internet Relay Chat reporting on the movements and activities of other players in this live Army exercise in a simulated set of Afghani villages. This excerpt illustrates the sort of gap that we are focusing on. Camera operator Heavy2 is reporting on an event involving a car 'of interest' in the Wakil village that he is observing.

Line	Time	Message
72	[19:51]	<Heavy2> black ford suv has entered wakil
73	[19:52]	<Heavy2> two people are dismounting

Table 1. Team chat excerpt from Empire Challenge 2010.

It is obvious to us where the people came from. In this section we lay out how we make it equally obvious to the C3 System.

4.1 The initial situation

Line 72 of the chat transcript, entered at 19:51 pm, is the first time that observer Heavy2 has typed anything for several minutes. This speaker shift has cleared the situation of any active habitats or facts, and moved their content to a passive store from which they can be reactivated when mentioned again. In this case, the "*black Ford SUV*" was already identified and designated as a 'vehicle of interest' earlier at 18:27, and at 18:50 there was the report "*three guys have gotten in to black ford suv at wakil.*" Not only is there a known individual to add to the situation (rather than building a new individual), but something is already known about it.[9]

```
SUV-1:  container.contents = collection(count > 3, type = person)
```

The discourse history established that The SUV is value of the **given** indexical variable. The value of the **new** variable is the fact that it has entered the village. This reintroduces the already known village in to the situation model, along with the fact of the event, but nothing else. The present location of the SUV is know (it is part of the minimal model), but nothing is known anything about its previous location except that it had one: "*where the SUV was before it entered Wakil.*"

From the text there is nothing else known about the SUV, not even whether it has stopped moving. But n the actual world of the observer, all of this is an esablished part of reality. It approached along a particular road at a particular angle to the viewer. The sun was shining and created shadow of a particular size. The buildings in Wakil are made of concrete and painted some color. All of this is true, but only what is actually given in the text is present in the situation. The rest is latent.

4.2 Expectations

In C3, texts are parsed incrementally word by word so as to get the greatest amount of leverage from the situation. From line 73, reported a minute after the report about the SUV, it first reads the word *two*. As a nominal premodifier, this deploys a peg and its packet establishes that there is a collection of size two, but that is all that is known at that moment: the rest of the text could refer to two of the windows on the SUV being opened, or two of its doors.

```
Peg(x):  collection(count = 2, type = x)
```

Upon reading *people*, the head of the NP, the peg is replaced by an individual representing a collection of two people, but again we know nothing more. There is an expectation, however. The people must have been somewhere before this even if we do not yet know where. Since some things, like the locations of the objects of discourse, are essential to understanding. (physical objects do not just appear in a puff of smoke), this information gap leads to an expectation that we will either be told the location or should assume one given the available evidence:

---

9. The expressions used in this section are purely notional for purposes of illustration. In C3's implementation their equivalents are configurations of typed objects linked by pointers and organized by indexical-variables bound by the situation object. We cannot describing their actual elements and organization in the space available.

```
people-2:  type = collection-2, physical-object.location = ?)
```

4.3 Composition

C3 then reads the verb group of line 73, "*are dismounting.*" It adds the packet for *dismount* to the situation and notes that this is an ongoing action.

```
dismount = transition.inprogress
   movement.from = high
   movement.to = low(ground)
   movement.actor = v:subject
```

From the syntactic construction, it knows that the collection of people supplies the obligatory argument to *dismount*: who is doing the action.

Dismount is a movement. Every instance of a movement comes with predicates for where its participants (the two people who are moving) were before the action and where they are after it. None of these values have been given explicitly, although a firm default for **dismount** is that the final location is the ground. (One dismounts from a horse or a piece of gymnastics equipment.)

To establish the value of their prior location (where they dismounted *from*), C3 uses what amounts to anaphoric reasoning: namely, what are the known locations given the present situation? This gives us the village and the SUV, but the SUV should be preferred because the thing one dismounts from must be close by (compare "*two people are walking up to it*") and the SUV is salient because it is the value the discourse **theme** indexical because is is a 'vehicle of interest':

```
during.before(dismount-1):
   people-2.physical-object.location = SUV-1
   dismount-1.movement.from = SUV-1
```

This binding furthermore has significant side effects. To be dismounting from the SUV presupposes that it is stopped, so C3 coerces the motion of the SUV in 72 to a 'stopped state.' (Compare secret service agents dismounting from the presidential limo during a motorcade.) If two people have *left* the SUV, qua container, then the number of people known to be in the vehicle (at least four) is reduced by two.

What has happened is that the introduction of the **dismount** to the situation initiated a limited inference process to identify the location the people dismounted from. Integrating the **dismount** with the established **enter** or the SUV provides a 'people-containing' location to the inferential search ('inside the SUV'). If there had not already been such a location in the current situation, the search would not go any further, and just posit that the location exists and wait for more information to come in, just as with our initial example of the Iraqi girl.

This example has illustrated some of the mechanics of C3's incremental analysis: Partial interpretations impose constraints on how they can be completed. Implicatures from inherited categories in the ontology (e.g. every physical object has a location) create expectations as they are incorporated into the situation, and can lead to constrained searches of the content of the situation that are organized by automatically-maintained indexical variables. We originally intended to extend this small model to the entire Empire Challenge chat corpus. However, we discovered that the inferential gap illustrated by this example is unique; the rest of the corpus can be understood with just a

literal analysis. Consequently we are shifting our future work to our original choice of topic, the inferentially rich domain of following route directions in hiking guides. This will let us develop vivid minimal simulation models and apply our extensive background in spatial and temporal ontologies (Pustejovsky, 2013a).

5. Related work

To the best of our knowledge, the first person to describe 'gap filling' was Herb Clark (1975) where he called this inferential process *bridging*. Bridging is "the construction of implicatures" in order to "bridge the gap from what [the listener] knows to what is intended" (pg. 170). In a logical framework, the process of adding implicatures in order to make sense of a text is usually treated as a form of *abduction*. This sort of defeasible reasoning has been studied at lenth by Hobbs (1993, 2011), who does language understanding by finding the least cost proof of the text's logical form.

We can also be said to be doing abduction in that we add implicatures to bridge the gaps in the text ('the men had been in the SUV'), however we do not formulate this as search over propositions and axioms by a theorem prover as Hobbs does. Instead, we take the psycholinguist evidence seriously and use an architecture that anticipates relevant inferences as latent variables and other kinds of inference that are introduced into the ongoing situation at virtually no cost and deployed only when a gap in the text triggers them.

In the 1970s there was work on frame-based language understanding, but either it formulated the problem in ways that could not be extended such as the anticipated questions approach of Charniak (1975), or it followed theoretical assumptions that have since been rejected: separating syntactic, semantic and pragmatic analysis into cascaded independent modules.

6. Claims

In this paper, we have presented a computational architecture for a new way to encode and exploit the knowledge and inferences that comprise a word's meaning. Our approach incorporates many ideas from the classical AI literature, but it also introduces innovations designed to increase efficiency and semantic transparency. These include:

- Treating situations computationally as the sum of the understanding of what has been said, along with what is implied and what might follow (Section 2.1).
- Organizing the meaning of words as 'packets' of model-level content along with overt and implicit predications (Section 2.2).
- Using a representation that enables constant-time access to the knowledge that is latent in the situation (Section 2.3).
- Providing functional landmarks to the content of a situation to permit one-step application of anaphoric-style inferential gaps (Section 2.5).
- Separating the linguistic realization from the statement of the axiomatic facts by defining predicates just once and stating wording constraints as restrictions on them (Section 2.6).

Clearly, there is much more to be fleshed out, and it is difficult to evaluate our proposal without more elaborate and extensive modeling. But the outline presented here suggests a specific way in which

people deploy their linguistic and general knowledge jointly to understand discourse, and we invite those who think that there is merit in our goal — to understand how people can use their knowledge for language as quickly and effortlessly as they walk or breathe — to an extended conversation about how this is possible.

Acknowledgements

References

Agre, P. E. (1988). *The dynamic structure of everyday life* (Technical Report). DTIC Document.

Agre, P. E., & Chapman, D. (1987). Pengi: An implementation of a theory of activity. *AAAI* (pp. 286–272).

Allen, J., Manshadi, M., Dzikovska, M., & Swift, M. (2007). Deep linguistic processing for spoken dialogue systems. *Proceedings of the Workshop on Deep Linguistic Processing* (pp. 49–56).

Bartlett, F. C. (1932). *Remembering*. Cambridge University Press.

Barwise, J., & Perry, J. (1983). *Situation semantics*. Stanford, CA: CSLI Publications.

Bergen, B., Chang, N., & Narayan, S. (2004). Simulated action in an embodied construction grammar. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 108–113).

Charniak, E. (1975). Organization and inference in a frame-like system of common sense knowledge. *Proceedings of the 1975 workshop on Theoretical issues in natural language processing* (pp. 42–51).

Charniak, E. (1983). Passing markers: A theory of contextual influence in language comprehension*. *Cognitive science*, *7*, 171–190.

Church, K. W., & Mercer, R. L. (1993). Introduction to the special issue on computational linguistics using large corpora. *Computational linguistics*, *19*, 1–24.

Clark, H. H. (1975). Bridging. *Proceedings of the 1975 workshop on Theoretical Issues in Natural Language Processing (TINLAP)* (pp. 169–174).

Clark, P., Harrison, P., & Thompson, J. (2003). A knowledge-driven approach to text meaning processing. *Proceedings of the HLT-NAACL 2003 Workshop on Text Meaning* (pp. 1–6).

Cottrell, G. W., & Small, S. L. (1983). A connectionist scheme for modelling word sense disambiguation. *Cognition & Brain Theory*.

Devlin, K. (2006). Situation theory and situation semantics. *Handbook of the History of Logic*, *7*, 601–664.

Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, *280*, 20–32.

Fillmore, C. J., & Baker, C. F. (2001). Frame semantics for text understanding. *Proceedings of WordNet and Other Lexical Resources Workshop*. Pittsburgh: NAACL.

Gabriel, R. P., White, J. L., & Bobrow, D. G. (1991). Clos: Integrating object-oriented and functional programming. *Communications of the ACM*, *34*, 29–38.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics volume 3: Speech acts*. Academic Press, New York.

Guha, R., & Lenat, D. (1993). Cyc: A midterm report. *Readings in knowledge acquisition and learning* (pp. 839–866).

Luperfoy, S. (1992). The representation of multimodal user interface dialogues using discourse pegs. *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics* (pp. 22–31).

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 522–523.

Matsuki, K., Chow, T., Hare, M., Elman, J., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 913–934.

McDonald, D. (1992). An efficient chart-based algorithm for partial-parsing of unrestricted texts. *Proceedings of the Third Conference on Applied Natural Language Processing* (pp. 193–200).

McDonald, D. (1996). The interplay of syntactic and semantic node labels in parsing. *Recent advances in parsing technology*, *1*, 295.

McDonald, D., & Pustejovsky, J. (1985). Tags as a grammatical formalism for generation. *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics* (pp. 94–103).

McDonald, D. D. (1994). Reversible nlp by linking the grammar to the knowledge base. In *Reversible grammar in natural language processing*, 257–291. Springer.

McDonald, D. D. (2000). Issues in the representation of real texts: The design of krisp. In L. M. Iwanska & S. C. Shapiro (Eds.), *Natural language processing and knowledge representation*, 77–110. MIT Press.

McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and linguistics compass*, *3*, 1417–1429.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision*, 211–277. McGraw-Hill.

Montazeri, N., & Hobbs, J. R. (2011). Elaborating a knowledge base for deep lexical semantics. *Proceedings of Ninth International Workshop on Computational Semantics* (pp. 195–204).

Newll, A., & Simon, H. A. *Human problem solving", year = "1972", publisher = Prentis Hall*.

Pustejovsky, J. (1991). The syntax of event structure. *Cognition*, *41*, 47–81.

Pustejovsky, J. (1995). *The generative lexicon*. MIT Press.

Pustejovsky, J. (2013a). Dynamic event structure and habitat theory. *Proceedings of GL2013* (pp. 1–20).

Pustejovsky, J. (2013b). Type theory and lexical decomposition. In *Advances in generative lexicon theory*, 9–38. Springer Netherlands.

Pustejovsky, J., & Boguraev, B. (1993). Lexical knowledge representation and natural language processing. *Artificial Intelligence*, *63*, 193–223.

Schank, R., & Abelson, R. (1977). *Sciprts plans goals and understanding*. Lawrence Erlbaum.

Schubert, L. K., & Hwang, C. H. (2000). Episodic logic meets little red riding hood. In L. M. Iwanska & S. C. Shapiro (Eds.), *Natural language processing and knowledge representation*, 111–174. MIT Press.

Small, S. L., Cottrell, G. W., & Tanenhaus, M. K. (Eds.). (1988). *Lexical ambiguity resolution*. Morgan Kauffmann.

Speer, N., Reynolds, J., Swallow, K., & Zacks, J. (2009). Reading stories activates neural representations of visual and motor experiences. *Psychological Science*, *20*, 989–999.

Speer, R., Havasi, C., & Lieberman, H. (2008). Analogyspace: Reducing the dimensionality of common sense knowledge. *AAAI* (pp. 548–553).

Swinney, D. A., & Hakes, D. T. (1976). Effects of prior context upon lexical access during sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *15*, 681–689.

Tanenhaus, M. K., Burgess, C., & Seidenberg, M. (1988). Is multiple access an artifact of backward priming. *Lexical Ambiguity Resolution: Perspectives from Psycholinguistics, Neuropsychology, and Artificial Intelligence. San Mateo, CA: Morgan Kaufmann*, 311–330.

Van Durme, B., Michalak, P., & Schubert, L. K. (2009). Deriving generalized knowledge from corpora using wordnet abstraction. *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 808–816).

von Humboldt, W. (1876). *Über die verschiedenheit des menschlichen sprachbaues und ihren einfluß auf die geistige entwickelung des menschengeschlechts*, Vol. 1. S. Calvary.