Towards an Encyclopedia of Compositional Semantics: Documenting the Interface of the English Resource Grammar

Dan Flickinger♣, Emily M. Bender◇, and Stephan Oepen♠♡

♣ Stanford University, Center for the Study of Language and Information
◇ University of Washington, Department of Linguistics
♠ University of Oslo, Department of Informatics
♡ Potsdam University, Department of Linguistics

danf@stanford.edu, ebender@uw.edu, oe@ifi.uio.no

Abstract

We motivate and describe the design and development of an emerging encyclopedia of compositional semantics, pursuing three objectives. We first seek to compile a comprehensive catalogue of interoperable semantic analyses—i.e. a precise characterization of meaning representations for a broad range of common semantic phenomena. Second, we operationalize the discovery of semantic phenomena and their definition in terms of what we call their semantic fingerprint, a formal account of the building blocks of meaning representation involved and their configuration. Third, we ground our work in a carefully constructed semantic test suite of minimal exemplars for each phenomenon, along with a 'target' fingerprint that enables automated regression testing. We work towards these objectives by codifying and documenting the body of knowledge that has been constructed in a long-term collaborative effort, the development of the LinGO English Resource Grammar. Documentation of its semantic interface is a prerequisite to use by non-experts of the grammar and the analyses it produces, but this effort also advances our own understanding of relevant interactions among phenomena, as well as of areas for future work in the grammar.

Keywords: Semantic phenomena; grammar engineering; English Resource Grammar

1. Motivation

In this work, we pursue three objectives, one theoretical, one methodological, and one practical. First, we seek to compile a *comprehensive catalogue* of semantic analyses—i.e. a precise characterization of meaning representations for a broad range of common semantic phenomena—that are *mutually compatible*. Second, we operationalize the *discovery and definition* of semantic phenomena in terms of what we call their *semantic fingerprint*, a formal account of the building blocks (of meaning representation) involved and their configuration. Third, we ground our work in a carefully constructed *semantic test suite* of minimal exemplars for each phenomenon, paired with a free-text discussion of relevant properties, as well as a 'target' fingerprint that enables automated *regression testing* in a grammar engineering context.

We work towards these objectives by codifying and documenting the body of knowledge that has been constructed in a long-term collaborative effort, the development of the LinGO English Resource Grammar (ERG; see §2.). As a general-purpose, linguistically motivated resource, the ERG is designed to provide a declarative, bidirectional mapping between form and meaning, i.e. between natural language utterances and associated logical-form semantic analyses. Reflecting some twenty-five person years of design and development to date, the grammar encompasses a large body of design decisions about the syntax–semantics interface, all implemented jointly in one consistent formal system and interacting in subtle ways.[1] Documentation of the semantic interface of the ERG is a prerequisite to use

of the resource by non-experts, but this effort also advances our own understanding of relevant interactions among phenomena, as well as of areas for future work in the grammar.

2. Background

The ERG is an implementation of the grammatical theory of Head-driven Phrase Structure Grammar (HPSG; Pollard & Sag, 1994) for English, i.e. a computational grammar that can be used for parsing and generation. Development of the ERG started in 1993, building conceptually on earlier work on unification-based grammar engineering for English at Hewlett Packard Laboratories (Gawron et al., 1982). The ERG has continuously evolved through a series of R&D projects (and two commercial applications) and today allows the grammatical analysis of running text across domains and genres, using an inventory of 222 syntactic rules, and 84 lexical rules for inflection, derivation, and punctuation. The hand-built ERG lexicon of some 38,000 lemmata, organized in a rich hierarchy of lexical types, aims for complete coverage of function words and open-class words with 'non-standard' syntactic properties (e.g. argument structure). Built-in support for light-weight named entity recognition and an unknown word mechanism typically enable the grammar to derive complete syntactico-semantic analyses for $85-95$ percent of all utterances in standard corpora, including newspaper text, the English Wikipedia, or biomedical research literature (Flickinger et al., 2012, 2010; Adolphs et al., 2008).

Through what is called the LinGO Redwoods Treebank (Oepen et al., 2004), for each release of the ERG, a selection of development corpora is manually annotated (via discriminant-based treebanking) to identify the analysis from among the alternatives provided by the grammar that matches the most plausible reading of the item in context (or with the observation that no appropriate such analysis is

---

[1] We particularly acknowledge the sustained contributions of Ann Copestake and Ivan Sag over the years to the designs of the semantic analyses implemented in the ERG; Copestake also had a key role in developing a closely related semantic test suite, which our exemplar construction (in §8.) extends in various dimensions.

available). In mid-2013, the current version of Redwoods encompasses gold-standard ERG analyses for 85,400 utterances (~1.5 million tokens) of running text from half a dozen different genres and domains, including the first 22 sections of the venerable Wall Street Journal (WSJ) text in the Penn Treebank (PTB; Marcus et al., 1993).

A core element of the design of HPSG ensures that descriptions of signs (words or phrases) include not only syntactic but also semantic constraints. Consistent with the underlying grammatical theory, ERG analyses maintain semantic compositionality, with the meaning of a given phrase consisting of the meanings of its components, together with any semantic contribution of the syntactic or lexical rule used to construct the phrase. These analyses also preserve monotonicity, so that any semantic contribution of a sign remains present in any larger containing constituent.

3. MRS Fundamentals

Meaning representation and composition in the ERG builds on the formal framework of Minimal Recursion Semantics (MRS; Copestake et al., 2005), a description language for logical-form semantics that affords scope underspecification. Our focus in this work, however, is on representation (and composition) more than on logic, and we will not dwell on details such as scope underspecification or interpretation.

We illustrate key properties of MRS with Figure 1, the semantic analysis derived by the ERG for (1):

(1)   The song was later covered by Harry Nilsson.

Basic Building Blocks   The representation in Figure 1 formally is a triple $\langle T, R, C \rangle$, with T being the (global) top handle, R a set of *elementary predications*, and C a set of handle constraints. Each elementary predication (EP) is comprised of a *predicate* (an identifier of the relation) and a set of *role–argument* pairs, where the roles draw from a small, fixed inventory of role lables: ARG0, ..., ARG4 for 'regular' arguments; RSTR and BODY on generalized quantifiers; and a few others in more specialized constructions). Role values (i.e. arguments) are variables. Predicates can be optionally parameterized by a constant argument, and these parameters are noted as a CARG pseudo-role. Finally, handle constraints are an element of the MRS mechanics of scope underspecification, expressing a binary relation between two handles. The ERG limits itself to only one type of handle constraint, called $=_q$, representing handle equality modulo quantifier insertion.

Variable Types   There are three types of variables in MRS, *event(ualitie)s* (of type e_i), *instances* (of type x_j), and labels or handles (of type h_k; shown as prefixes on predicates in Figure 1). Of these, the latter serve a formalism-internal role, to identify groups of conjoined predications, designate scopal arguments, and facilitate scope underspecification; assuming a suitable variant of predicate logic as the object language, MRS handles will not map onto logical variables. Eventualities and instances, on the other hand, prototypically correspond to the semantics of verbal and nominal expressions, respectively, i.e.

$\langle h_1,$
$\quad h_4{:}\_the\_q\langle 0{:}3\rangle(\text{ARG0 } x_6, \text{RSTR } h_7, \text{BODY } h_5),$
$\quad h_8{:}\_song\_n\_of\langle 4{:}8\rangle(\text{ARG0 } x_6, \text{ARG1 } i_9),$
$\quad h_2{:}\_later\_a\_1\langle 13{:}18\rangle(\text{ARG0 } e_{10}, \text{ARG1 } e_3),$
$\quad h_2{:}\_cover\_v\_1\langle 19{:}26\rangle(\text{ARG0 } e_3, \text{ARG1 } x_{11}, \text{ARG2 } x_6),$
$\quad h_{13}{:}proper\_q\langle 30{:}44\rangle(\text{ARG0 } x_{11}, \text{RSTR } h_{14}, \text{BODY } h_{15}),$
$\quad h_{16}{:}compound\langle 30{:}44\rangle(\text{ARG0 } e_{18}, \text{ARG1 } x_{11}, \text{ARG2 } x_{17}),$
$\quad h_{19}{:}proper\_q\langle 30{:}35\rangle(\text{ARG0 } x_{17}, \text{RSTR } h_{20}, \text{BODY } h_{21}),$
$\quad h_{22}{:}named\langle 30{:}35\rangle(\text{ARG0 } x_{17}, \text{CARG } Harry),$
$\quad h_{16}{:}named\langle 36{:}44\rangle(\text{ARG0 } x_{11}, \text{CARG } Nilsson)$
$\{ h_{20} =_q h_{22}, h_{14} =_q h_{16}, h_7 =_q h_8, h_1 =_q h_2 \} \rangle$

Figure 1: MRS meaning representation for example (1).

eventuality variables denote states or activities, while instance variables will typically correspond to (referential or abstract) entities.

In addition to these most specific variable types, there are underspecifications as follows: *i* (for *individual*) is a supertype of eventualities and instances; *p* (the half-way mark in the alphabet between *h* and *x*) is a supertype of labels and instances; and *u* (for *unspecific* or maybe unbound) is a supertype of all of the above.

Although not shown in Figure 1, non-label variables can be 'refined' with what we call *variable properties*, e.g. TENSE or SF (sentence force) on eventualities, and NUM(ber) or PERS(on) on instances. Variable properties record semantic reflexes of morpho-syntactic information and range over a fixed inventory of possible values, organized in a multiple-inheritance hierarchy to allow underspecification.

Construction Semantics   While most EPs correspond to particular lexical items, EPs can also be contributed by grammar rules. In Figure 1 for example, the two-place compound relation provides the bracketing of the complex proper name *Harry Nilsson*. Syntax, however, does not necessarily determine the exact internal structure of complex nominals, e.g. the (likely) interpretation of *Harry* as a first name, in this case. We return to the analysis of underspecified compound structures and to the notion of construction semantics in §4. and §6. below, respectively.

Argument Identification   In the ERG, all elementary predications have an ARG0 role, providing the *intrinsic argument* of a relation, e.g. the instance variable introduced by the semantics of a nominal expression, or the eventuality corresponding to the semantics of a verbal expression in a (Neo-)Davidsonian representation.[2] For what is quite likely a blend of linguistic and technical reasons, the ERG also introduces intrinsic variables in the semantics of, among others, adjectives, adverbs, and prepositions—eventualities in all three cases.

In addition to ARG0, most EPs will take additional arguments, labeled ARG1, ARG2 etc. Note that these labels are not intended to evoke consistent roles across different predicates, though within a predicate their interpretation is consistent. In particular, observe how role assignment is normalized at the level of semantics: the mapping of syntactic functions to semantic arguments is reversed in the passive

---

[2]The distinguished nature of the ARG0 role, contrasting with other arguments, led Oepen & Lønning (2006) to the term *distinguished variable*, whereas Copestake (2009) coined the name *characteristic variable* for the same general notion.

construction, but in ERG ouputs the passive and active variants receive identical semantics: the instance variable associated with *Harry Nilsson*, x_{11}, is the ARG1 ('deep subject') of the _cover_v_1 predication—as would be the case with other diathesis alternations analyzed by the grammar, e.g. the dative shift in *Kim gave Sandy a book.* vs. *Kim gave a book to Sandy.* For a more in-depth motivation of these design decisions in ERG semantics, please see the discussion of 'slacker semantics' by Copestake (2009).

4. Semantic Fingerprints

We characterize individual semantic phenomena in terms of their 'fingerprints', i.e. those pieces of a full MRS which uniquely identify the phenomenon. We utilize a compact template language for fingerprints that closely resembles the query language defined by Kouylekov & Oepen (2014) for searching large MRS collections. In our fingerprint language, irrelevant roles, handles, and other information like character spans can be left unspecified, and we further allow wild-carding over predicate symbols, role labels, and variable properties. As an example, (2) shows the semantic fingerprint for basic nominal compounding, as in e.g. *garden dog*:

$$(2) \quad \left\langle \begin{array}{|l} h_0:\text{compound}(\text{ARG1 } x_1, \text{ARG2 } x_2), \\ h_0:\_(\text{ARG0 } x_1), \\ \_(\text{ARG0 } x_2) \\ \{\,\} \end{array} \right\rangle$$

This phenomenon is characterized by the appearance of the underspecified two-place compound relation, linking together another two EPs in the configuration indicated by the shared label h_0 (of the compound head and the two-place modifier relation) and the shared instance variables x_1 and x_2. Comparing (2) to Figure 1, the fingerprint match operationalizes the generalization that the ERG analysis of complex names constitutes an instance of the semantic phenomenon of (nominal) compounding.

Besides their use in making explicit (and operational) how we define individual semantic phenomena in the space of meaning representations available from the ERG, we anticipate that the fingerprint template language and automated comparison of fingerprints to full MRSs will have engineering utility in various contexts. For example, the approach to searching very large collections of MRSs developed by Kouylekov & Oepen (2014) can be combined with our inventory of semantic phenomena and associated fingerprints to retrieve and aggregate instances of these phenomena in naturally occurring text, e.g. ERG treebanks or treecaches. Such statistics could serve, for example, to approximate relative frequencies of phenomena, quantitative patterns of interaction with other phenomena, and also an estimate of what proportion of an average ERG analysis is covered (i.e. matched by fingerprints) by explicit documentation.

5. A Near-Complete Example

To give an idea of what the documentation for each phenomenon looks like, Figure 2 shows about two thirds of the current page for compounding. This documentation is grounded in a collection of test suite examples (see §8. below), and each page instantiates a templatic structure, with a pre-defined set and ordering of section headers. In addition to the ones shown in Figure 2, the template comprises sections for (a) additional *Motivating Examples*, building on richer context or more specialized vocabulary than is generally used in the test suite, or contrasting with negative examples; (b) a discussion of (non-trivial) *Interactions* with other semantic phenomena; and (c) comparatively free-form reflections on the phenomenon, the development status of its semantic analysis, or relevant aspects of composition and the syntax–semantics interface.

There is of course great variation in the complexity and internal sub-division of different phenomena and, thus, the appropriate breadth and depth of documentation. Our running compounding example ranges somewhere in the middle of this (intuitive) scale. At the level of meaning representation, the underspecified semantic analysis assigned by the ERG is comparatively simple and very regular. At the same time, there is non-trivial syntactic variation (involving for example common nouns, proper names, and titles, as well as non-nominal heads) in this construction. Likewise, the same two-place compound relation is introduced (twice) in a little-studied but not infrequent construction that Bender et al. (2011) term N-ed, as for example in *rabbit-eared dog*. Here, despite the seemingly verbal inflection on *eared*, the analysis is in terms of nested intersective modification between three instance variables, i.e. analogous to a paraphase using two prepositions: *dog with ears like a rabbit*. Thus, in this example, there is a comparatively high degree of normalization in mapping between a range of syntactic structures and an underspecified meaning representation.

6. Phenomena Discovery

The first step in the creation of the encyclopedia is the development of an inventory of analyses of phenomena to document. The delineation of linguistic phenomena is a notoriously difficult problem, as linguistic categories tend to have central prototypes and fuzzy boundaries, and furthermore can be classifed at verious levels (e.g. coordination subsumes NP coordination, VP coordination, S coordination, etc. as well as coordination of unlike constituents). However, as our task is to document the analyses developed within the ERG, we do not need an *a priori* listing of phenomena but rather can work from the grammar to discover how it equates and distinguishes different sentence types. This process is *data-driven* on two levels: On a conceptually prior level, the analyses in the ERG were developed in the context of and to account for naturally occurring linguistic data. Building on that, we extract parts of the ERG encoding semantic analyses and classify them according to common properties in order to arrive at our notion of 'phenomena'. We thus developed a discovery procedures that starts from what we call *grammar entities*: phrase structure rules, lexical rules, and lexical entries. This discovery procedure can be understood as a means to classify the syntactic building blocks in the grammar according to their semantic effects; the set of categories that emerge from that classification are the first pass at our set of phenomena.

Our procedure starts from the identification of grammar entities which are likely to contribute to the composition of

```
    The garden dog barked.
    The Abrams picture arrived.
    Kim Browne arrived.
    Professor Browne arrived.
    The tobacco-happy dog barked.
    The state and local dogs barked.
```

↪Linguistic Characterization

Compounding comprises a variety of (semantic) head–modifier structures that can often be paraphrased using overt prepositions. But in instances of compounding there is a syntactic construction contributing an underspecified two-place relation. *Medicine delivery*, for example, could be paraphrased as *delivery of medicine*, whereas *home delivery* might be paraphrased as *delivery at home*. The phenomenon is characterized by the underspecified compound relation, whose shape and use are parallel to the semantics contributed by regular prepositions. While prototypical heads in compounds are nominals (introducing an instance), there can be exceptions: *tobacco-happy* is analyzed in terms of an event head. Compounding can interact with coordinate structures (e.g. in *state and local dogs*), when modifiers of the same head are conjoined.
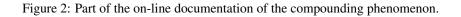
↪MRS Fingerprints

The phenomenon can be sub-divided according to the (semantic) types of the head and modifier. The most basic instances, arguably, involve non-quantified or quantified modifiers to a nominal head, e.g. *garden dog* and *Abrams picture*, respectively.

```
    h0:compound(ARG1 x1, ARG2 x2)
    h0:_(ARG0 x1)
    _(ARG0 x2)
```

With event heads, the structure is very similar, except for the type of the ARG1 in the compound relation, e.g. for *tobacco-happy*:

```
    h0:compound(ARG1 e1, ARG2 x1)
    h0:_(ARG0 e1)
    _(ARG0 x1)
```

Figure 2: Part of the on-line documentation of the compounding phenomenon.

semantic analyses that go 'beyond the basics'. For phrase structure and lexical rules, we identify all instances which contribute predications of their own. For the lexical entries, we work from what the ERG calls lexical types, i.e. a hand-built collection of around a thousand fine-grained lexical categories. In our search for non-basic semantic phenomena that manifest themselves in lexical types, we included those that (a) contribute more than one predication, i.e. have lexically decomposed semantics; or (b) take at least one scopal argument.

Each grammar entity (or type) activated by these discovery heuristics is then associated with a signature, for the time being just its (multi-)set of semantic predicates. We then create rough clusters based on shared signatures, clustering only within broad entity classes (phrase structure rules, lexical rules, or lexical types). Each such cluster presents a candidate semantic phenomenon. The ERG (in its 1212 release) distinguishes 222 phrase structure and 84 lexical rules; of these, 135 and 27, respectively, were activated as potentially interesting grammar entities, and the corresponding semantic signatures form 39 and 18 distinct clusters, respectively.

The next step in the procedure is to manually inspect the clusters to determine whether they are internally consistent and whether any should be combined into larger clusters. We found that this step was greatly facilitated by looking at examples of actual sentences (from running text) that illustrate the functioning of the grammar entities in question. To find these examples, we indexed the items in the Redwoods Treebank by their usage of relevant grammar entities and then extracted the three shortest examples for each grammar entity. An example showing the results of this procedure for two rules in the compounding cluster is shown in Figure 3.

```
J-N_N-ED_C: compound udef_q

|Is Dumbo four-legged?| 1 3 fracas
|She is good-natured.| 1 3 rtc000
|These cookies are star-shaped.| 1 4 rtc000
Total instances in all corpora: 73

N-HDN_TTL-CPD_C: compound udef_q

|Bus stop/parking at Borgund Stave church.| 1 6 rondane
|Fondsbu tourist lodge was opened in 1993.| 1 7 jh2
|Olavsbu self-service cabin was opened in 1952.| 1 7 jh2
Total instances in all corpora: 31
```

Figure 3: Examples extracted for compounding cluster.

Working from the candidate clusters derived for the phrase structure rules and lexical rules and their associated examples, we derived an initial set of about two dozen high-level semantic phenomena to validate and document. These range from the mundane (including, for example, the semantics associated with complex literal numbers or foreign-language expressions, fragmentary utterances, or parentheticals) to the intricate, as for example the semantics of measure phrases, partitives, conditionals, or coordination. Some of our candidate phenomena identified by MRS signatures straddle the semantics–pragmatics divide, in the sense of providing what to us seems like an MRS-based encoding of quasi-semantic or para-semantic information, e.g. the foregrounding effects of topicalization or *it* clefts.

While this procedure proved efficient and effective for classifying phenomena implemented in terms of semantically contentful phrase structure rules and lexical rules, the set of lexical types that meet our heuristics above is still too large and at the same time the heuristics are likely masking

lexical types that should be explored. In future work, we plan to refine the discovery procedure as applied to lexical types in order to develop a more comprehensive inventory of semantic phenomena to document. In the meantime, we have identified an initial set of a dozen or so phenomena whose semantic analyses are encoded in lexical types, including such elements as cardinal adjectives, control relations, pro-verbs (e.g. 'do so'), and identity copulae.

7. Towards an Inventory of Semantic Phenomena

Our emerging encyclopedia is currently implemented as a set of wiki pages,[3] populated with an initial list of phenomena that emerged from the discovery procedure described above, including 22 that are tied to syntactic or lexical rules. These encompass frequently occurring phenomena such as coordination, compounds, and nominalization, as well as less frequently encountered phenomena such as vocatives or instrumental relative clauses. At present, there are separate pages now in place for more than a third of these phenomena, providing for each a brief linguistic characterization, some motivating examples, the MRS fingerprint(s), discussion of interactions with other phenomena, and a set of open questions about the chosen MRS representation, to assist in further refinements of either the documentation or the grammar.

Already, the process of recording these descriptions has in several instances led to the identification of unnecessary differentiation in the MRSs assigned to closely related variants, pointing the way to desirable improvements to the ERG toward more uniformity. For example, the implementation of the rules for some of the types of compounds introduced slightly different names for the two-place relation added by the constructions, and in the absence of any observed benefit for these varient names, we arrived at a beneficial normalizing of the predicate name across the range of compounding phenemena.

We expect to have completed the documentation of these 22 rule-anchored phenomena within a matter of weeks, and will then turn our attention to the refinement and documentation of the list of phenomena anchored in lexical types. The existing wiki page already includes a list of some 16 such phenomena, including comparative and superlative adjectives, possessives, control predicates, and verb particle constructions. This inventory will grow as we improve our discovery procedure for identifying distinctive MRS fingerprints among the 1000 lexical types employed in the ERG.

8. Semantic Test Suite

As we examine each semantic phenomenon in turn, we identify phenomenon-internal semantic variation and construct an emerging test suite of minimal sentences, to include at least one canonical exemplar of each subphenomenon. Test suite entries are constructed by simplification of running-text examples found through the discovery procedure sketched above, eliminating irrelevant syntactic complexity and minimizing vocabulary variation. The test suite is parsed with the ERG and disambiguated

using the standard Redwoods methodology. Careful study of the MRS that the ERG assigns in the intended analysis of each exemplar informs our documentation effort, and is the basis for a four-way classification of our degree of confidence in the actual analysis. In some cases, current analyses reflect a careful design process (e.g. building on supporting literature or revisions of earlier attempts); in other cases, there may be known minor deficiencies; and for yet another group of semantic phenomena, current ERG analyses may be mere placeholders ('sewing things together' somehow, without a deep commitment to the analysis); or an analysis may simply be broken, i.e. formally not well-formed or otherwise indefensible.

The current test suite consists of some 65 sentences, each exhibiting some distinctive single semantic phenomenon which derives some aspect of its composition from a syntactic construction or lexical rule which is not semantically transparent. For conditionals, for example, the test suite includes the pair *If Abrams arrived, Browne barked.* and *Had Abrams arrived, Browne would have barked.* These exemplars yield structurally isomorphic semantics but differ in the specific predicate encoding the (lexical or structural) conditional. However, syntactic variants that result from merely re-ordering the two clauses are not represented in our test suite, as they are semantically equivalent.

More generally, the test suite may for a given phenomenon include multiple examples for which the ERG assigns the same MRS, in order to confirm that differing paths of composition (due to different syntactic structures) do in fact arrive at the same MRS when desired. For example, we may include for the partitive construction not only *All of the dogs bark* but also *All the dogs bark* to ensure that the grammar produces the same MRS for both variants. The test suite can also include multiple MRS-equivalent examples for constructions where that semantic equivalence may be unexpected or not obvious. For example, the ERG assigns nearly identical MRS representations for the following two examples illustrating possessive constructions in English:

(3) My old friends arrived.
     Old friends of mine arrived.

Here the ERG is rather carefully designed to assign (nearly) the same MRS to both the lexical and the periphrastic possessives, and the inclusion of both in the test suite helps to ensure the preservation of that equivalence as the grammar develops further.

In ongoing work, we are manually adding the semantic fingerprint to each test suite sentence for which its analysis is considered stable. This pairing enables both human inspection of semantic phenomena and automated regression testing, i.e. confirming the integrity of semantic analyses for new versions of the ERG. Since we employ the Redwoods treebanking method for recording the intended semantic analyses in the test suite, we can maintain the accuracy of these representations with each revision of the ERG, by automatically re-applying the disambiguating decisions to the freshly computed parse forest for each sentence. If a change to the grammar inadvertently affects one or more of the recorded fingerprints, such regression can be automatically flagged, thus enabling efficient diagnosis and

---

correction of the grammar. Once we expand our documentation to include lexicalized non-basic semantic effects, this test suite will grow significantly in size.

9. Related Work

Where most theoretical work on compositional semantics focuses on fundamentals (e.g. the representation of predicate–argument structure or of scopal operators), handling compositional semantics in a broad-coverage grammar also requires working out representations of many different phenomena which are encountered in naturally occurring language, working out which syntactically disparate phenomena should in fact map to the same semantic representations, and finally making sure the analyses of various individual phenomena are interoperable with each other. It is in this sense that the analyses implemented in the ERG have been developed in a data-driven fashion and as such it is of interest to compare the collection of analyses we document in the encyclopedia with other collections of descriptions of semantic phenomena which have been developed on the basis of exploration of naturally occurring text.

The only such collections we have been able to identify are the annotation guidelines for large-scale semantic annoation projects, including the English PropBank Annotation Guidelines (Bonial et al., 2012), the FrameNet annotation guidelines (Ruppenhofer et al., 2010), and the Abstract Meaning Representation (AMR) Specification (Banarescu et al., 2014). One fundamental difference between semantic annotation guidelines and our encyclopedia of semantic analyses is that the guidelines were developed to guide humans to consistently annotate the phenomena of interest in the same way, whereas we are documenting the analyses that are programmatically produced by the ERG. These analyses are manually developed, but their progammatic application raises the bar for both interoperability of the analyses with each other and comprehensiveness: for the grammar to work in both parsing and generation, every component of every analyzed sentence must either make a specific contribution to the semantic representation or be explicitly designated as semantically empty. On the other hand, while the underlying grammar must meet this bar of comprehensiveness, the encyclopedia itself is a work in progress with only partial coverage over the phenomena analyzed.

PropBank and FrameNet in particular are only focused on annotating specific parts of semantic representations (the identification of a subset of predicates and their arguments and modifiers). For example, PropBank does not address the issue of noun-noun compounds (though FrameNet does, looking to identify specific roles for the non-head if the head noun is 'frame-evoking'); neither PropBank nor FrameNet specify representations for conditionals. AMR on the other hand does strive for spanning analyses, though unlike PropBank, FrameNet and the semantic representations produced by the ERG, AMR incorporates a design decision to create representations that are not closely tied to the syntactic structure of the strings being annotated. This abstraction from the actual utterance among other things leaves room to not annotate the contribution of particular elements.

Nonetheless, all of these resources represent discussions of design decisions about semantic representations, and one scientific benefit of the existence of the documentation is the possiblity for detailed comparison where there is overlap. One example of a phenomenon that is addressed by all the resources is infinitival purpose clauses ('quasi-modal infinitives', in our resource).

10. Outlook & Future Work

We believe the methodology sketched above offers a good balance of automation and expert analysis. We are currently working through our initial inventory of semantic phenomena represented in the ERG, in each instance reviewing current analyses and relevant literature, to then provide a high-level textual discussion of the phenomenon (and any subphenomena) as well as its formal fingerprint. For maximum accessibility, we prepare an on-line, open-source repository of documentation and supporting data,[4] which we hope will grow over time into a community-supported 'encyclopedia' of ERG semantic analyses. As with other long-term opensource projects of non-trivial complexity, we anticipate that an organizational challenge may arise in balancing community involvement vs. correctness and authority of the available information. With a relatively small core group of developers and a large remaining body of implicit shared knowledge, we believe the 'open collaborative' authoring model of a resource like Wikipedia may not be the ideal setup, and we will probably rather look for methodological inspiration in a (much larger) project like the Linux kernel.

However, our ambitions transcend improvement and documentation of one grammar. There is growing interest in parsing into formal representations of meaning, often using the label 'semantic parsing' (e.g. Yahya et al., 2012, Cai & Yates, 2013, and Kwiatkowski et al., 2013, inter alios). Some of this work frames the task in terms of mapping from surface strings to application-specific representations of meaning. Our approach instead involves crafting general-purpose, application-independent semantic representations which strive to capture all and only the cues to *speaker meaning* encoded as the conventional *sentence meaning* of a given utterance.

Formal semantic representations as a generic interface to parsing are subject to constraints from multiple disciplines: (computational) linguistics capitalizes on representational adequacy and compositional compatibility with syntax and morphology; formal semantics traditionally has its focus on aspects of mathematical logic, truth conditions, and support for inference; application development, finally, emphasizes more practical requirements—for example that representations are sufficiently detailed, yet easy to comprehend,

---

[4]See `http://moin.delph-in.net/ErgSemantics`. However, we already experience that the combination of a set of wiki pages coupled with maintenance of the test suite in a revision control system leaves something to be desired in terms of technological infrastructure. To better integrate data and documentation, we are currently preparing a formal, structured file format (inspired by the TeXinfo markup language of the GNU project) from which machine- and human-readable views on the information will be auto-generated.

and stable over time. Reconciling such different points of view remains a major challenge and, we believe, a barrier to wider uptake of broad-coverage parsing into general-purpose semantic target representations. Once there is sufficient critical mass of manually reviewed, documented, exemplified, and fingerprinted ERG analyses, we will invite others to pair our semantic test suite with alternative views on phenomena characterization and semantic analysis. Ideally, this community exercise will result in cross-framework, parallel, aligned semantic annotations and will thus further the scholarly discourse about balancing requirements on computational semantics along the above three dimensions.

References

Adolphs, P., Oepen, S., Callmeier, U., Crysmann, B., Flickinger, D., & Kiefer, B. (2008). Some fine points of hybrid natural language parsing. In *Proceedings of the 6th International Conference on Language Resources and Evaluation.* Marrakech, Morocco.

Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., ... Schneider, N. (2014). *Abstract Meaning Representation (AMR) 1.1 specification.* Retrieved from http://www.isi.edu/~ulf/amr/help/amr-guidelines.pdf (Version of February 11, 2014)

Bender, E. M., Flickinger, D., Oepen, S., & Zhang, Y. (2011). Parser evaluation over local and non-local deep dependencies in a large corpus. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (p. 397–408). Edinburgh, Scotland, UK.

Bonial, C., Hwang, J., Bonn, J., Conger, K., Babko-Malaya, O., & Palmer, M. (2012). *English PropBank annotation guidelines.* Retrieved from https://verbs.colorado.edu/propbank/EPB-Annotation-Guidelines.pdf

Cai, Q., & Yates, A. (2013). Large-scale semantic parsing via schema matching and lexicon extension. In *Proceedings of the 51th Meeting of the Association for Computational Linguistics* (p. 423–433). Sofia, Bulgaria: Association for Computational Linguistics.

Copestake, A. (2009). Slacker semantics. Why superficiality, dependency and avoidance of commitment can be the right way to go. In *Proceedings of the 12th Meeting of the European Chapter of the Association for Computational Linguistics* (p. 1–9). Athens, Greece.

Copestake, A., Flickinger, D., Pollard, C., & Sag, I. A. (2005). Minimal Recursion Semantics. An introduction. *Research on Language and Computation*, *3*(4), 281–332.

Flickinger, D., Oepen, S., & Ytrestøl, G. (2010). WikiWoods. Syntacto-semantic annotation for English Wikipedia. In *Proceedings of the 7th International Conference on Language Resources and Evaluation.* Malta.

Flickinger, D., Zhang, Y., & Kordoni, V. (2012). Deep-Bank. A dynamically annotated treebank of the Wall Street Journal. In *Proceedings of the 11th International Workshop on Treebanks and Linguistic Theories* (p. 85–96). Lisbon, Portugal: Edições Colibri.

Gawron, J. M., King, J., Lamping, J., Loebner, E., Paulson, E. A., Pullum, G. K., ... Wasow, T. (1982). Processing English with a Generalized Phrase Structure Grammar. In *Proceedings of the 20th Meeting of the Association for Computational Linguistics* (p. 74–81). Toronto.

Kouylekov, M., & Oepen, S. (2014). Semantic technologies for querying linguistic annotations. An experiment focusing on graph-structured data. In *Proceedings of the 9th International Conference on Language Resources and Evaluation.* Reykjavik, Iceland.

Kwiatkowski, T., Choi, E., Artzi, Y., & Zettlemoyer, L. (2013). Scaling semantic parsers with on-the-fly ontology matching. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (p. 1545–1556). Seattle, WA, USA: Association for Computational Linguistics.

Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpora of English: The Penn Treebank. *Computational Linguistics*, *19*, 313–330.

Oepen, S., Flickinger, D., Toutanova, K., & Manning, C. D. (2004). LinGO Redwoods. A rich and dynamic treebank for HPSG. *Research on Language and Computation*, *2*(4), 575–596.

Oepen, S., & Lønning, J. T. (2006). Discriminant-based MRS banking. In *Proceedings of the 5th International Conference on Language Resources and Evaluation* (p. 1250–1255). Genoa, Italy.

Pollard, C., & Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. Chicago, USA: The University of Chicago Press.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, R., Christopher, & Scheffczyk, J. (2010). *FrameNet II: Extended theory and practice.* Retrieved from https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf

Yahya, M., Berberich, K., Elbassuoni, S., Ramanath, M., Tresp, V., & Weikum, G. (2012). Natural language questions for the web of data. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Conference on Natural Language Learning* (p. 379–390). Jeju Island, Korea: Association for Computational Linguistics.