A Type-Theoretical system for the FraCaS test suite: Grammatical Framework meets Coq

Jean-Philippe Bernardy and Stergios Chatzikyriakidis

September 20, 2017

centre for
linguistic theory
and studies in probability

Brief summary of the talk

- Present a type-theoretical framework for formal semantics leveraging two well-studied tools
  - Grammatical Framework (GF, Ranta 2004, 2011)
  - Coq
- Providing a compositional resource semantics for GF
- Evaluation on the FraCaS test suite
- State-of-the-art results

CLASP centre for linguistic theory and studies in probability

Structure the talk

- Brief intro to the systems used, GF and Coq
- Presenting the FraCoq system
  - We concentrate on the most linguistically relevant features and also the features relevant for the FraCaS
- Evaluation against the FraCaS test suite
  - Some brief remarks about the FraCaS and NLI platforms
    - Results and comparison with previous logical approaches
    - The issue of automation
- Conclusions and Future work

CLASP centre for linguistic theory and studies in probability

Background: Grammatical Framework (GF)

- Functional programming language for writing multi-lingual grammars
- Involves an abstract syntax, comprised of:
  - A number of syntactic categories
  - A number of syntactic construction functions, which provide the means to compose basic syntactic categories into more complex ones
    - $AdjCN{:}AP \to CN \to CN$ (appending an adjectival phrase to a common noun and obtaining a new common noun)
- GF comes with a library of mappings from abstract syntax to concrete
  - These mappings can be inverted by GF, thus offering parsers from natural text into abstract syntax
  - We use the parse trees constructed by Ljunglöf (2012) thereby avoiding any syntactic ambiguity (GF FraCaS treebank).

CLASP centre for linguistic theory and studies in probability

Background: Type-Theoretical Semantics

- We use the type of logics that have been traditionally dubbed as constructive
  - Initiated by the work of Martin-Löf (1971, 1984)
  - In linguistics this types of logics go back to Ranta's seminal work (Ranta 1994) or even earlier to Sundholm (1988)
    - More recent approaches can be found as well. Please see Chatzikyriakidis and Luo (2017) for a collection of papers on constructive type theories for NL semantics

centre for linguistic theory and studies in probability

Background: Type Theoretical Semantics

- Main features of MTTs
  - Type many-sortedness.
  - Dependent sum and product types
    - Σ-types, often written $\sum_{x:A} B[x]$ and which have product types $A \times B$ as a special case when B does not depend on x.
    - Dependent product, Π-types, often written $(\prod_{x:A} B[x])$, and which have arrow-types $A \to B$ as a special case
    - They generalize universal quantification and function types and they offer type polymorphism
  - Proof-theoretical specification and support for effective reasoning.
    - Most powerful proof assistants implement MTTs (e.g. Coq, Agda)

CLASP centre for linguistic theory and studies in probability

Background: Coq

- Proof assistant based on the calculus of inductive constructions (CiC, Coquand and Paulin-Mohring 1988)
  - Arguably one of the leading proof assistants
    - a proof of the four-color theorem (Gonthier 2008)
    - a proof of the odd order theorem (Gonthier et al. 2013)
    - developing CompCert, a formally verified compiler for C (Leroy 2013)
    - One of the assistants used in the Univalent Foundations project (Homotopy Type Theory, The Univalent Foundations Program 2013)

Background: Coq

- Important features used
  - Π types
    - in Coq: $\prod_{x:A} B[x]$ is written `forall (x:A), B` or (simply `A` $\rightarrow$ `B` when `B` does not depend on `x`)
  - Record types
    - Generalization of Σ-types and are encoded as (trivial) inductive types with a single constructor.
    - $\Sigma x{:}A.B(x)$ can be expressed as a dependent record type in Coq:

```
Record AB:Type:=mkAB{x:> A;P:B x}
```

CLASP centre for linguistic theory and studies in probability

The FraCoq system

- We use Ljünglof's FraCaS treebank and take these trees to their semantic counterparts
- The structure of the semantic representation
    1. Every GF syntactic category C is mapped to a Coq *Set*, noted $[\![C]\!]$.
    2. GF Functional types are mapped compositionally : $[\![A \rightarrow B]\!] = [\![A]\!] \rightarrow [\![B]\!]$
    3. Every GF syntactic construction function $f{:}X$ is mapped to a function $[\![f]\!]$ such that $[\![f]\!] : [\![X]\!]$.
    4. GF function applications are mapped compositionally: $[\![t(u)]\!] = [\![t]\!]([\![u]\!])$.

CLASP centre for linguistic theory and studies in probability

The FraCoq system

- Sentences
  - We interpret sentences as propositions: $[\![S]\!] = Prop$.
  - To verify that P entails H, we prove the proposition $[\![P]\!] \rightarrow [\![H]\!]$.

  ```
  Definition S := Prop.
  ```
- Common Nouns
  - Predicates over an abstract object type

  ```
  Parameter object : Set.
  Definition CN := object->Prop.
  ```

CLASP centre for linguistic theory and studies in probability

The FraCoq system

- Verb phrases
  - Parameterize over the *noun* of the subject (using Π types)

```
Definition VP := forall (subjectClass : CN)
object -> Prop.
```

CLASP centre for linguistic theory and studies in probability

The FraCoq system

- Adjectives
  - Functions from cn to cn (predicates to predicates)

    ```
    Definition A := CN -> CN.
    ```

  - Different classes of adjectives are captured using coercions (subtyping). All special classes of adjectives are subtypes of A.

    ```
    Definition IntersectiveA := object -> Prop.
    Definition wkIntersectiveA : IntersectiveA -> A
    := fun a cn (x:object) => a x /\ cn x.
    Coercion wkIntersectiveA : IntersectiveA >-> A.
    ```

  - Provision is made for intersective, subsective, privative and non-committal adjectives

CLASP centre for linguistic theory and studies in probability

The FraCoq system

- Adverbs
  - Similar method to adjectives but instead the modification is on verbal predicates
  - The adverb cases in the FraCaS are all veridical and covariant.
  - We define such a subclass *VeridicalAdv* and declare it as a coercion *Adv*
    - ⋆ Adverbs of type *VeridicalAdv* are also of type *Adv*

```
Parameter on_time_Adv : VeridicalAdv .
```

CLASP centre for linguistic theory and studies in probability

The FraCoq system

- Noun Phrases and Predeterminers
  - A clean definition of NPs as functions from predicates to truth values will not work
    - Problem with GF's abstract syntax: existence of pre-determiners which include cases like *most*, *all* among others and are defined as functions from NPs to NPs
    - In general, the category includes elements that are naturally interpreted as GQs
    - Solution: Remember the components of NPs (number, quantifier and common noun)
    - Pre-determiners then are able to substistute the quantifier part with the appropriate quantifier
    - This has to be done, otherwise pre-determiners introduce a dummy indefinite in these cases

CLASP centre for linguistic theory and studies in probability

The FraCoq system

- Generalized Quantifiers
  - They turn a number and a common noun into a noun-phrase (which we call *NP*0).

    ```
    Definition Quant := Num -> CN -> NP0.
    ```
  - Some quantifiers ignore the number and are given usual definitions (e.g. *some* or *all*), whereas others make essential use of number (e.g. *at most*)
    - ★ In the latter case, the function `CARD`, a context-dependent abstract function which turns a predicate into a natural number is used to get the correct semantics

CLASP centre for linguistic theory and studies in probability

The FraCoq system

- The definite article
  - Checks for plural noun phrases
    - ★ If found, then universal quantification
    - ★ If not, it looks up the object of discourse in an abstract *environment*, which is a function which turns a common noun into an object

```
Definition DefArt:Quant:= fun (num : Num) (P:CN)=> fun Q:VP
=> match num with plural => (forall x, P x -> Q P x)
 /\ Q P (environment P) /\ P (environment P) |
_ => Q P (environment P) /\ P (environment P) end.
```

CLASP centre for linguistic theory and studies in probability

Evaluation: The FraCaS test suite

- A test suite for NLI
  - 346 NLI examples in the form of one or more premises followed by a question along with an answer to that question
  - Three potential answers
    - ★ YES: The declarative sentence formed out of the question follows from the premises
    - ★ NO: The declarative sentence does not follow from the premises
    - ★ UNK: The declarative sentence neither follows nor does not follow fro the premises

CLASP centre for linguistic theory and studies in probability

Evaluation: The FraCaS test suite

(1) A Swede won the Nobel Prize.
Every Swede is Scandinavian.
Did a Scandinavian win the Nobel prize? [Yes, FraCas 049]

(2) No delegate finished the report on time..
Did any Scandinavian delegate finish the report on time? [No, FraCaS 070]

(3) A Scandinavian won the Nobel Prize.
Every Swede is Scandinavian.
Did a Swede win the Nobel prize? [UNK, FraCaS 065]

CLASP centre for linguistic theory and studies in probability

Evaluation: The FraCaS test suite

- The FraCaS has considerable weaknesses
  - Small size
  - Artificial nature of the examples
- However, it covers a lot of phenomena associated with NLI
- It is still a very good suite to test logical approaches

CLASP centre for linguistic theory and studies in probability

Evaluation

- Evaluation against 5 sections of the FraCaS
  - Total of 174 examples
  - Excluded sections where a lot of context-dependency has to be taken into consideration (e.g. the section on ellipsis)
- YES: a proof can be constructed from the premises to the hypothesis
- NO: a proof of the negated hypothesis can be constructed
- UNK: otherwise

CLASP centre for linguistic theory and studies in probability

Evaluation

- The following table presents the results (Ours) as well as a comparison with the approach in Mineshima et al. (MINE, 2015), Bos (Nut, 2008) and Abzianidze (Langpro, 2015)

	Section	# examples	Ours	MINE	Nut	Langpro
1	Quantifiers	75	.96	.77	.53	.93 (44)
2	Plurals	33	.76	.67	.52	.73 (24)
3	Adjectives	22	.95	.68	.32	.73 (12)
4	Comparatives	31	.56	.48	.45	-
5	Attitudes	13	.85	.77	.46	.92 (9)
6	Total	174 (181)	0.83	0.69	0.50	0.85

- Our approach outerperforms Mineshema et al. by 13 percentage points.
- The approach by Abzianidze has an accuracy of 0.85 without involving the comparative section. If this section is taken out, our system's accuracy rises to 0.88

CLASP centre for linguistic theory and studies in probability

Error Analysis

- Improvement over earlier approaches. Still, there were a couple of difficulties
  - Comparatives: cases that needed one to provide adequate semantics for *more* but also to take care of ellipsis

    (4) ITEL won more orders than APCOM.
    ITEL won some orders.
    Did ITEL win some orders? [Yes, FraCaS 233]

  - Definite Plurals: Universal reading was captured. Cases of existential readings were not

    (5) The inhabitants of Cambridge voted for a Labour MP.
    very inhabitant of Cambridge voted for a Labour MP.
    Did every inhabitant of Cambridge vote for a Labour MP?
    [UNK, FraCaS 094]

CLASP centre for linguistic theory and studies in probability

Automation

- So far, our proofs are not automated
  - A couple of steps (usually very few) to reach a proof
  - Earlier approaches using Coq (e.g. Chatzikyriakidis 2014 and Mineshima et al. 2015) use Coq's tactical language LTac to define automated macros of actions
    - ★ This is not difficult to do in our case as well
    - ★ Just go through all the proof tactics or observe the tactics that are used in the proofs to create a macro that will automate the proofs
    - ★ The question remains: can that macro of tactics generalize outside the suite?
    - ★ Answer: only to a limited extent, i.e. when exactly the same set of tactics yields a proof
    - ★ For this reason, we have not automated proof search to obtain the results presented in this paper, even though this can be done easily

CLASP centre for linguistic theory and studies in probability

Automation

- Automating would also make an unprincipled use of higher-order logic (HOL)
  - No algorithm which can decide if a proposition has a proof or not
    - We must use heuristics both to search for proofs and to decide when to give up searching
- Most problems have either obvious proofs or obviously lack a proof (fortunately)
  - Due to its heuristic nature the proof search necessarily contains a human component
    - Problematic to make a statement about the suitability of FraCoq outside FraCas
    - Small dataset and lack of separation between a development and a test set does not help the situation either
    - Related shortcoming: specialized semantics for specific lexical entries

CLASP centre for linguistic theory and studies in probability

Future Work

- Address the issue of automation
  - Define a decidable fragment of the logic and only work within such fragment
    - Possible to concisely characterize how the approach generalises
  - Train a neural network on a body of freely available proofs on the net and see whether it can generalize to automatically provide the proof tactics for the cases interested
- Improvement at the GF level: make the abstract syntax more compatible with compositional semantics
  - For example, do something about problematic syntactic categories like pre-determiners or cases where the syntax makes it impossible to recover elliptical fragments
- Extend into the whole suite (first attempt to do anaphora using monads on the way!)

CLASP centre for linguistic theory and studies in probability

Conclusions

- We have connected two well-defined systems based on type-theory
  - GF and Coq
  - Providing resource semantics for GF
- The issue of generalization remains a shortcoming
  - It is possible to achieve very precise semantics for specific domains
    - Our system outerperforms previous logical systems w.r.t accuracy
- Useful in performing inference tasks on controlled natural language domains
- Hybrid NLI systems

CLASP centre for linguistic theory and studies in probability

Conclusions

- The system can be found here: https://github.com/GU-CLASP/FraCoq

CLASP centre for linguistic theory and studies in probability