A CCG-based Compositional Semantics and Inference System for Comparatives

Izumi Haruta
Ochanomizu University
haruta.izumi@is.ocha.ac.jp

Koji Mineshima
Ochanomizu University
mineshima.koji@ocha.ac.jp

Daisuke Bekki
Ochanomizu University
bekki@is.ocha.ac.jp

Abstract

Comparative constructions play an important role in natural language inference. However, attempts to study semantic representations and logical inferences for comparatives from the computational perspective are not well developed, due to the complexity of their syntactic structures and inference patterns. In this study, using a framework based on Combinatory Categorial Grammar (CCG), we present a compositional semantics that maps various comparative constructions in English to semantic representations, and introduce an inference system that effectively handles logical inference with comparatives, including those involving numeral adjectives, antonyms, and quantification. We evaluate the performance of our system on the FraCaS test suite and show that the system can handle a variety of complex logical inferences with comparatives.

1 Introduction

Gradability is a pervasive phenomenon in natural language and plays an important role in natural language understanding. Gradable expressions can be characterized in terms of the notion of *degree*. Consider the following examples:

(1) a. My car is more *expensive* than yours.
b. My car is *expensive*.

The sentence (1a), in which the comparative form of the gradable adjective *expensive* is used, compares the price of two cars, making it a comparison between degrees. The sentence (1b), which contains the positive form of the adjective, can be regarded as a construction that compares the price of the car to some implicitly given degree (i.e., price).

In formal semantics, many in-depth analyses use a semantics of gradable expressions that relies on the notion of degree (Cresswell, 1976; Kennedy, 1997; Heim, 2000; Lassiter, 2017, among others). Despite this, meaning representations and inferences for gradable expressions have not been well developed from the perspective of computational semantics in previous research (Pulman, 2007). Indeed, a number of logic-based inference systems have been proposed for the task of Recognizing Textual Entailment (RTE), a task to determine whether a set of premises entails a given hypothesis (Bos, 2008; MacCartney and Manning, 2008; Mineshima et al., 2015; Abzianidze, 2016; Bernardy and Chatzikyriakidis, 2017). However, these logic-based systems have performed relatively poorly on inferences with gradable constructions, such as those collected in the FraCaS test suite (Cooper et al., 1994), a standard benchmark dataset for evaluating logic-based RTE systems (see §5 for details).

There are at least two obstacles to developing a comprehensive computational analysis of gradable constructions. First, the syntax of gradable constructions is diverse, as shown in (2):

(2) a. Ann is tall. (Positive)
b. Ann is taller than Bob. (Phrasal)
c. Ann is taller than Bob is. (Clausal)
d. Ann is as tall as Bob. (Equative)
e. Ann is $2''$ taller than Bob. (Differential)

In the examples above, (2c) is a clausal comparative

33rd Pacific Asia Conference on Language, Information and Computation (PACLIC 33), pages 47-56, Hakodate, Japan, September 13-15, 2019

in which *tall* is missing from the subordinate *than*-clause. (2e) is an example of a differential comparative in which a measure phrase, $2''$ (*2 inches*), appears. The diversity of syntactic structures makes it difficult to provide a compositional semantics for comparatives in a computational setting.

Second, gradable constructions give rise to various inference patterns that require logically complicated steps. For instance, consider (3):

(3) P_1: Mary is taller than 4 feet.
P_2: Harry is shorter than 4 feet.
H: Mary is taller than Harry.

To logically derive H from P_1 and P_2, one has to assign the proper meaning representations to each sentence, and those representations include numeral expressions (*4 feet*), antonyms (*short/tall*), and their interaction with comparative constructions.

For these reasons, gradable constructions pose an important challenge to logic-based approaches to RTE, serving as a testbed to act as a bridge between formal semantics and computational semantics.

In this paper, we provide (i) a compositional semantics to map various gradable constructions in English to semantic representations (SRs) and (ii) an inference system that derives logical inference with gradable constructions in an effective way. We will mainly focus on gradable adjectives and their comparative forms as representatives of gradable expressions, leaving the treatment of other gradable constructions such as verbs and adverbs to future work.

We use Combinatory Categorial Grammar (CCG) (Steedman, 2000) as a syntactic component of our system and the so-called *A-not-A analysis* (Seuren, 1973; Klein, 1980, 1982; Schwarzschild, 2008) to provide semantic representations for comparatives (§2, §3). We use ccg2lambda (Martínez-Gómez et al., 2016) to implement compositional semantics to map CCG derivation trees to SRs. We introduce an axiomatic system COMP for inferences with comparatives in typed logic with equality and arithmetic operations (§4). We use a state-of-the-art prover to implement the COMP system. We evaluate our system[1] on the two sections of the FraCaS test suite (ADJECTIVE and COMPARATIVE) and show that it can handle various complex inferences with gradable adjectives and comparatives.

[1]All code is available at:
https://github.com/izumi-h/fracas-comparatives_adjectives

2 Background

2.1 Comparatives in degree-based semantics

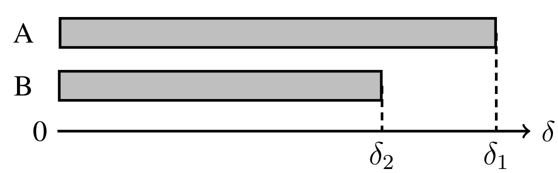
To analyze gradable adjectives, we use the two-place predicate of entities and degrees as developed in degree-based semantics (Klein, 1982; Kennedy, 1997; Heim, 2000; Schwarzschild, 2008). For instance, the sentence *Ann is 6 feet tall* is analyzed as $\mathbf{tall}(\mathbf{Ann}, 6 \text{ feet})$, where $\mathbf{tall}(x, \delta)$ is read as "x is (at least) as tall as degree δ".[2]

In degree-based semantics, there are at least two types of analyses for comparatives. Consider (4), a schematic example for a comparative construction.

(4) A is taller than B is.

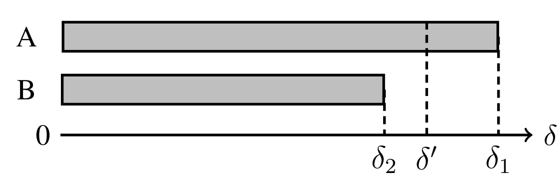
The first approach is based on the maximality operator (Stechow, 1984; Heim, 2000). Using the maximality operator ($\max$) as illustrated in (5), the sentence (4) is analyzed as a statement asserting that the maximum degree δ_1 of A's tallness is greater than the maximum degree δ_2 of B's tallness.

(5) $\max(\lambda\delta.\mathbf{tall}(A, \delta)) > \max(\lambda\delta.\mathbf{tall}(B, \delta))$



The other approach is the A-not-A analysis (Seuren, 1973; Klein, 1980, 1982; Schwarzschild, 2008). In this type of analysis, (4) is treated as stating that there exists a degree δ' of tallness that A satisfies but B does not, as shown in (6).

(6) $\exists\delta\,(\mathbf{tall}(A, \delta) \wedge \neg\,\mathbf{tall}(B, \delta))$



[2]For simplicity, we do not consider the internal structure of a measure phrase like *6 feet*. For an explanation of why $\mathbf{tall}(x, \delta)$ is not treated as "x is *exactly* as tall as δ", see, e.g., Klein (1982).

Table 1: Semantic representations of basic comparative constructions

Type	Example	SR
Increasing Comparatives	Mary is taller than Harry.	$\exists\delta\,(\mathbf{tall}(\mathbf{m},\delta) \wedge \neg\,\mathbf{tall}(\mathbf{h},\delta))$
Decreasing Comparatives	Mary is less tall than Harry.	$\exists\delta(\neg\mathbf{tall}(\mathbf{m},\delta) \wedge \mathbf{tall}(\mathbf{h},\delta))$
Equatives	Mary is as tall as Harry.	$\forall\delta(\mathbf{tall}(\mathbf{h},\delta) \rightarrow \mathbf{tall}(\mathbf{m},\delta))$

Table 2: Semantic representations of complex comparative constructions

Type	Example	SR
Subdeletion Comparatives	Mary is taller than the bed is long.	$\exists\delta(\mathbf{tall}(\mathbf{m},\delta) \wedge \neg\,\mathbf{long}(\mathbf{the}(\mathbf{bed}),\delta))$
Measure phrase comparatives	Mary is taller than 4 feet.	$\exists\delta(\mathbf{tall}(\mathbf{m},\delta) \wedge (\delta > 4'))$
Differential Comparatives	Mary is 2 inches taller than Harry.	$\forall\delta(\mathbf{tall}(\mathbf{h},\delta) \rightarrow \mathbf{tall}(\mathbf{m},\delta + 2''))$
Negative Adjectives	Mary is shorter than Harry.	$\exists\delta\,(\mathbf{short}(\mathbf{m},\delta) \wedge \neg\,\mathbf{short}(\mathbf{h},\delta))$

Although the two analyses are related as illustrated in the figures (5) and (6), we can say that the A-not-A analysis is less complicated and easier to handle than the maximality-based analysis from a computational perspective, mainly because it only involves constructions in first-order logic (FOL).[3] We thus adopt the A-not-A analysis and extend it to various types of comparative constructions for which inference is efficient in our system.

2.2 Basic syntactic assumptions

There are two approaches to the syntactic analysis of comparative constructions. The first is the *ellipsis* approach (e.g. Kennedy, 1997), in which phrasal comparatives such as (2b), are derived from the corresponding clausal comparatives, such as (2c). The other is the *direct* approach (e.g. Hendriks, 1995), which treats phrasal and clausal comparatives independently and does not derive one from the other. An argument against the ellipsis approach is that it has difficulties in accounting for coordination such as that in (7) (Hendriks, 1995).

(7) a. Someone at the party drank more vodka than wine.
    b. Someone at the party drank more vodka than someone at the party drank wine.

Here, (7a), a phrasal comparative with an existential NP *someone*, does not have the same meaning as the corresponding clausal comparative (7b); the person who drank vodka and the one who drank wine do not have to be the same person in (7b), whereas they must be the same person in (7a).[4] In this study, we adopt the direct approach and use CCG to formalize the syntactic component of our system.

3 Framework

3.1 Semantic representations

Table 1 shows the SRs for basic constructions of comparatives under the A-not-A analysis we adopt. Using this standard analysis, we also provide SRs for more complex constructions, including subdeletion, measure phrases, and negative adjectives. Table 2 summarizes the SRs for these constructions.

Some remarks are in order about how our system handles various linguistic phenomena related to gradable adjectives and comparatives.

Antonym and negative adjectives *Short* is the antonym of *tall*, which is represented as $\mathbf{short}(x,\delta)$ and has the meaning "the height of x is less than or equal to δ". Thus, we distinguish between the monotonicity property of positive adjectives such as *tall* and *fast* and that of negative adjectives such as *short* and *slow*. For positive adjectives, if $\mathbf{tall}(x,\delta)$ is true, then x satisfies all heights below δ; by contrast, for negative adjectives, if $\mathbf{short}(x,\delta)$ is true, then x satisfies all the heights above δ.

In general, for a positive adjective F^+ and a negative adjective F^-, (8a) and (8b) hold, respectively.

(8) $\forall\delta_1\forall\delta_2 : \delta_1 > \delta_2 \rightarrow$
    a. $\forall x(F^+(x,\delta_1) \rightarrow F^+(x,\delta_2))$
    b. $\forall x(F^-(x,\delta_2) \rightarrow F^-(x,\delta_1))$

[3] See van Rooij (2008) for a more detailed comparison of the two approaches.

[4] See Hendriks (1995) and Kubota and Levine (2015) for other arguments against the ellipsis approach.

Positive form and comparison class As mentioned in §1, the positive form of an adjective is regarded as involving comparison to some threshold that can be inferred from the context of the utterance. We write $\theta_F(A)$ to denote the contextually specified threshold for a predicate F given a set A, which is called COMPARISON CLASS (Klein, 1982). When a comparison class is implicit, as in (9a) and (10a), we use the universal set U as a default comparison class[5]; we typically abbreviate $\theta_F(U)$ as θ_F. Thus, (9a) is represented as (9b), which means that the height of Mary is more than or equal to the threshold θ_{tall}. Similarly, the SR of (10a) is (10b), which means that the height of Mary is less than or equal to the threshold θ_{short}.

(9) a. Mary is tall.
b. $\mathbf{tall}(\mathbf{m}, \theta_{\text{tall}})$

(10) a. Mary is short.
b. $\mathbf{short}(\mathbf{m}, \theta_{\text{short}})$

A threshold can be explicitly constrained by an NP modified by a gradable adjective. Thus, (11a) can be interpreted as (11b), relative to an explicit comparison class, namely, the sets of animals.[6]

(11) a. Mickey is a small animal. (FraCaS-204)
b. $\mathbf{small}(\mathbf{m}, \theta_{\mathbf{small}}(\mathbf{animal})) \wedge \mathbf{animal}(\mathbf{m})$

Numerical adjectives We represent a numerical adjective such as *ten* in *ten orders* by the predicate $\mathbf{many}(x, n)$, with the meaning that the cardinality of x is at least n, where n is a positive integer (Hackl, 2000). For example, *ten orders* is analyzed as $\lambda x.(\mathbf{order}(x) \wedge \mathbf{many}(x, 10))$. The following shows the SRs of some typical sentences involving numerical adjectives.

(12) a. Mary won ten orders.
b. $\exists x(\mathbf{order}(x) \wedge \mathbf{won}(\mathbf{m}, x) \wedge \mathbf{many}(x, 10))$

(13) a. Mary won many orders.
b. $\exists\delta\exists x(\mathbf{order}(x) \wedge \mathbf{won}(\mathbf{m}, x) \wedge \mathbf{many}(x, \delta) \wedge (\theta_{\text{many}} < \delta))$

(14) a. Mary won more orders than Harry.
b. $\exists\delta(\exists x(\mathbf{order}(x) \wedge \mathbf{won}(\mathbf{m}, x) \wedge \mathbf{many}(x, \delta)) \wedge \neg\exists y(\mathbf{order}(y) \wedge \mathbf{won}(\mathbf{h}, y) \wedge \mathbf{many}(y, \delta)))$

3.2 Compositional semantics in CCG

Here we give an overview of how to compositionally derive the SRs for comparative constructions in the framework of CCG (Steedman, 2000). In the CCG-style compositional semantics, each lexical item is assigned both a syntactic category and an SR (represented as a λ-term). In this study, we newly introduce the syntactic category D for degree and assign $S\backslash NP\backslash D$ to gradable adjectives. For instance, the adjective *tall* has the category $S\backslash NP\backslash D$ and the corresponding SR is $\lambda\delta.\lambda x.\mathbf{tall}(x, \delta)$.

Table 3 lists the lexical entries for representative lexical items used in the proposed system. We abbreviate the CCG category $S\backslash NP\backslash D$ for adjectives as AP and $S/(S\backslash NP)$ (a type-raised NP) as $NP^{\uparrow}$.[7]

The suffix *-er* for comparatives such as *taller* is categorized into four types: clausal and phrasal comparatives (-er$_{\text{simp}}$), subdeletion comparatives (-er$_{\text{sub}}$), measure phrase comparatives (-er$_{\text{mea}}$), and differential comparatives (-er$_{\text{diff}}$). We assume that equatives are constructed from as$_{\text{simp}}$ and as$_{\text{cl}}$; for instance, the equative sentence in Table 1 corresponds to *Mary is as$_{simp}$ tall as$_{cl}$ Harry*. For measure phrase comparatives, such as *Mary is taller than 4 feet*, we use than$_{\text{deg}}$; and for comparatives with numerals, such as (14a), we use more$_{\text{simp}}$.

On the basis of these lexical entries, we can compositionally map various comparative constructions to suitable SRs. Some example derivation trees for comparative constructions are shown in Figure 1 and 2. An advantage of using CCG as a syntactic theory is that the *function composition* rule ($>$B) can be used for phrasal comparatives such as that in Figure 1, where the VP *is tall* is missing from the subordinate *than*-clause. For positive forms, we use the empty element *pos* of category $S\backslash NP/(S\backslash NP\backslash D)$, as shown in Figure 2.[8]

[5]In this case, we do not consider the context-sensitivity of the implicit comparison class. See Narisawa et al. (2013) for work on this topic in computational linguistics.

[6]Here and henceforth, when an example appears in the FraCaS dataset, we refer to the ID of the sentence in the dataset.

[7]We also abbreviate $\lambda X_1. \ldots \lambda X_n.M$ as $\lambda X_1 \ldots X_n.M$.

[8]Note that the role played by the empty element *pos* here can be replaced by imposing a unary type-shift rule from $S\backslash NP\backslash D$ to $S\backslash NP$.

Table 3: Lexical entries in CCG-style compositional semantics

PF	CCG categories	SR
tall	AP	$\lambda\delta x.\mathbf{tall}(x,\delta)$
Mary	NP	$\mathbf{mary}$
is	$S\backslash NP/(S\backslash NP)$	id
$4'$	D	$4'$
$\text{than}_{\text{simp}}$	S/S	id
than_{deg}	D/D	id
than_{gq}	$S\backslash NP\backslash(S\backslash NP/NP^{\uparrow})/NP^{\uparrow}$	$\lambda QWx.Q(\lambda y.W(\lambda P.P(y))(x))$
pos	$S\backslash NP/AP$	$\lambda A.A(\theta_A)$
-er_{simp}	$S\backslash NP/NP^{\uparrow}\backslash AP$	$\lambda AQx.\exists\delta(A(\delta)(x)\wedge\neg Q(A(\delta)))$
-er_{sub}	$S\backslash NP/(S\backslash D)\backslash AP$	$\lambda AKx.\exists\delta(A(\delta)(x)\wedge\neg K(\delta))$
-er_{mea}	$S\backslash NP/D\backslash AP$	$\lambda A\delta' x.\exists\delta(A(\delta)(x)\wedge(\delta>\delta'))$
-er_{diff}	$S\backslash NP/NP^{\uparrow}\backslash D\backslash AP$	$\lambda A\delta' Qx.\forall\delta(Q(A(\delta))\rightarrow A(\delta+\delta')(x))$
as_{simp}	$S\backslash NP/NP^{\uparrow}/AP$	$\lambda AQx.\forall\delta(Q(A(\delta))\rightarrow A(\delta)(x))$
as_{cl}	S/S	id
more_{num}	$S\backslash NP/NP^{\uparrow}\backslash(S\backslash NP/NP)/N$	$\lambda NGQz.\exists\delta(\exists x(N(x)\wedge G(\lambda P.P(x))(z)\wedge\mathbf{many}(x,\delta))$ $\wedge\neg\exists y(N(y)\wedge Q(G(\lambda P.P(y)))\wedge\mathbf{many}(y,\delta)))$
more_{is}	$S\backslash NP/NP^{\uparrow}\backslash(S\backslash NP/NP)/N/AP$	$\lambda ANGQz.\exists\delta(\exists x(N(x)\wedge G(\lambda P.P(x))(z)\wedge A(\delta)(x)))$ $\wedge\neg Q(\lambda y.(N(y)\wedge A(\delta)(x)))$
more_{has}	$S\backslash NP/NP^{\uparrow}\backslash(S\backslash NP/NP)/N/AP$	$\lambda ANGQz.\exists\delta(\exists x(N(x)\wedge G(\lambda P.P(x))(z)\wedge A(\delta)(x)))$ $\wedge\neg\exists y(N(y)\wedge Q(G(\lambda P.P(y)))\wedge A(\delta)(x))$

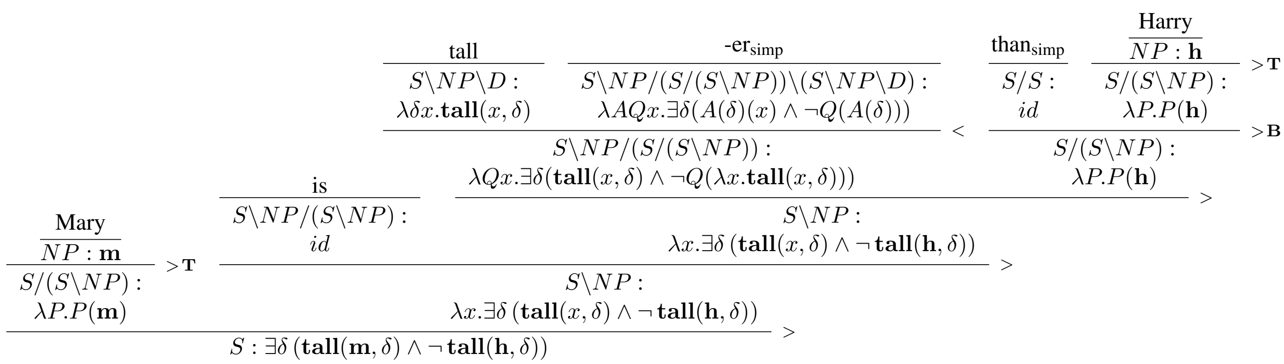


Figure 1: Derivation tree of *Mary is taller than Harry*

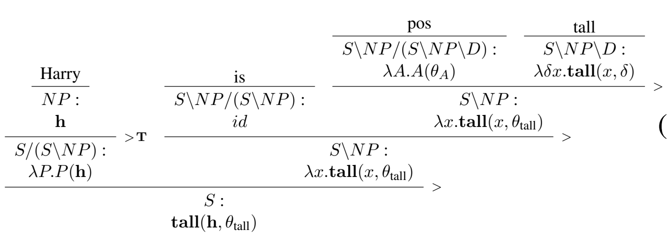


Figure 2: Derivation tree of *Harry is tall*

Quantification When determiners such as *all* or *some* appear in *than*-clauses, we need to consider the scope of the corresponding quantifiers (Larson, 1988). As examples, (15a) and (16a) are assigned the SRs in (15b) and (16b), respectively.

(15) a. Mary is taller than everyone.

b. $\forall y(\mathbf{person}(y)$
$\rightarrow \exists\delta\,(\mathbf{tall}(\mathbf{m},\delta)\wedge\neg\,\mathbf{tall}(y,\delta)))$

(16) a. Mary is taller than someone.

b. $\exists y(\mathbf{person}(y)$
$\wedge\,\exists\delta\,(\mathbf{tall}(\mathbf{m},\delta)\wedge\neg\,\mathbf{tall}(y,\delta)))$

Figure 3 shows a derivation tree for (15a). Here, *everyone* in *than*-clause takes scope over the degree quantification in the main clause. For this purpose, we use the lexical entry for than_{gq} in Table 3, which handles these cases of generalized quantifiers.

Conjunction and disjunction Conjunction (*and*) and disjunction (*or*) appearing in a *than*-clause show different behaviors in scope taking, as pointed out by Larson (1988). For instance, in (17a), the con-

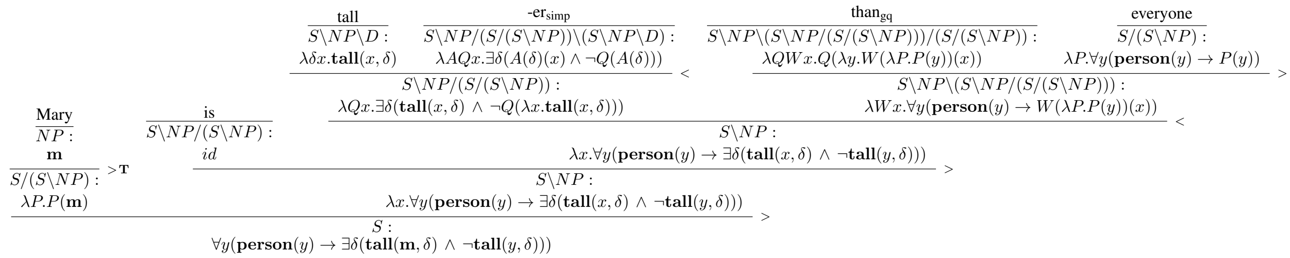


Figure 3: Derivation tree of *Mary is taller than everyone*

junction *and* takes wide scope over the main clause, whereas in (18a), the disjunction *or* can take narrow scope; thus, we can infer *Mary is taller than Harry* from both (17a) and (18a). These readings are represented as in (17b) and (18b), respectively.

(17) a. Mary is taller than Harry and Bob.

b. $\exists\delta\,(\mathbf{tall}(\mathbf{m},\delta) \wedge \neg\,\mathbf{tall}(\mathbf{h},\delta)) \wedge \exists\delta\,(\mathbf{tall}(\mathbf{m},\delta) \wedge \neg\,\mathbf{tall}(\mathbf{b},\delta))$

(18) a. Mary is taller than Harry or Bob.

b. $\exists\delta(\mathbf{tall}(\mathbf{m},\delta) \wedge \neg(\mathbf{tall}(\mathbf{h},\delta) \vee \mathbf{tall}(\mathbf{b},\delta)))$

The difference in scope for these sentences can be derived by using $\text{than}_{\text{simp}}$ and than_{gq}: $\text{than}_{\text{simp}}$ derives the narrow-scope reading (cf. the derivation tree in Figure 1) and than_{gq} derives the wide-scope reading (cf. the derivation tree in Figure 3).

Attributive comparatives The sentence *APCOM has a more important customer than ITEL* (FraCaS-244/245) can have two interpretations, i.e., (19a) and (20a), where the difference is in the verb of the *than*-clause.

(19) a. APCOM has a more important customer than ITEL is. (FraCaS-244)

b. $\exists\delta(\exists x(\mathbf{customer}(x) \wedge \mathbf{has}(\mathbf{a},x) \wedge \mathbf{important}(x,\delta)) \wedge \neg(\mathbf{customer}(\mathbf{i}) \wedge \mathbf{important}(\mathbf{i},\delta)))$

(20) a. APCOM has a more important customer than ITEL has. (FraCaS-245)

b. $\exists\delta(\exists x(\mathbf{customer}(x) \wedge \mathbf{has}(\mathbf{a},x) \wedge \mathbf{important}(x,\delta)) \wedge \neg\exists y(\mathbf{customer}(y) \wedge \mathbf{has}(\mathbf{i},y) \wedge \mathbf{important}(y,\delta)))$

We use more_{is} or more_{has} in Table 3 to give the compositional derivations of the SRs in (19b) and (20b), respectively.

4 Inferences with comparatives

We introduce an inference system COMP for logical reasoning with gradable adjectives and comparatives based on the SRs under the A-not-A analysis presented in §3. Table 4 lists some axioms of COMP for inferences with comparatives. Here, F is an arbitrary gradable predicate, F^+ a positive adjective, and F^- a negative adjective.[9]

$(\mathbf{CP})$ is the so-called Consistency Postulate (Klein, 1982), an axiom asserting that if there is a degree satisfied by x but not by y, then every degree satisfied by y is satisfied by x as well. By $(\mathbf{CP})$, we can derive the following inference rule.

$$(\mathbf{CP}\star)\ \frac{\exists\delta\,(\mathbf{F}(x,\delta) \wedge \neg\,\mathbf{F}(y,\delta))}{\forall e(\mathbf{F}(y,e) \to \mathbf{F}(x,e))}$$

Using this rule, the inference from *Mary is taller than Harry* and *Harry is tall* to *Mary is tall* can be derived as shown in Figure 4.

$$\dfrac{\dfrac{\dfrac{\exists\delta\,(\mathbf{tall}(\mathbf{m},\delta) \wedge \neg\,\mathbf{tall}(\mathbf{h},\delta))}{\forall e(\mathbf{tall}(\mathbf{h},e) \to \mathbf{tall}(\mathbf{m},e))}\,(\mathbf{CP}\star)}{\mathbf{tall}(\mathbf{h},\theta_{\text{tall}}) \to \mathbf{tall}(\mathbf{m},\theta_{\text{tall}})}\,(\forall\mathbf{E}) \quad \mathbf{tall}(\mathbf{h},\theta_{\text{tall}})}{\mathbf{tall}(\mathbf{m},\theta_{\text{tall}})}\,(\to\mathbf{E})$$

Figure 4: Example of a proof

$(\mathbf{Ax_1})$ and $(\mathbf{Ax_2})$ are axioms for positive and negative adjectives described in (8). The axioms from $(\mathbf{Ax_3})$ to $(\mathbf{Ax_6})$ formalize the entailment relations between antonym predicates. For instance, the inference of (3) mentioned in §1 is first mapped to the following SRs.

[9] We also use an axiom for privative adjectives such as *former*, drawn from Mineshima et al. (2015).

Table 4: Axioms of COMP

$(\mathbf{TH})$	$\theta_{F^+} > \theta_{F^-}$
$(\mathbf{CP})$	$\forall x \forall y (\exists \delta\, (F(x,\delta) \wedge \neg F(y,\delta)) \rightarrow (\forall e (F(y,e) \rightarrow F(x,e))))$
$(\mathbf{Ax_1})$	$\forall e \forall x (F^-(x,e) \leftrightarrow \forall \delta ((\delta \geq e) \rightarrow F^-(x,\delta)))$
$(\mathbf{Ax_2})$	$\forall e \forall x (F^+(x,e) \leftrightarrow \forall \delta ((\delta \leq e) \rightarrow F^+(x,\delta)))$
$(\mathbf{Ax_3})$	$\forall e \forall x (F^-(x,e) \leftrightarrow \forall \delta ((\delta > e) \rightarrow \neg F^+(x,\delta)))$
$(\mathbf{Ax_4})$	$\forall e \forall x (F^+(x,e) \leftrightarrow \forall \delta ((\delta < e) \rightarrow \neg F^-(x,\delta)))$
$(\mathbf{Ax_5})$	$\forall e \forall x (\neg F^-(x,e) \leftrightarrow \forall \delta ((\delta \leq e) \rightarrow F^+(x,\delta)))$
$(\mathbf{Ax_6})$	$\forall e \forall x (\neg F^+(x,e) \leftrightarrow \forall \delta ((\delta \geq e) \rightarrow F^-(x,\delta)))$

(21) P_1: $\exists \delta (\mathbf{tall}(\mathbf{m}, \delta) \wedge (\delta > 4'))$
P_2: $\exists \delta (\mathbf{short}(\mathbf{h}, \delta) \wedge (\delta < 4'))$
H: $\exists \delta\, (\mathbf{tall}(\mathbf{m}, \delta) \wedge \neg\, \mathbf{tall}(\mathbf{h}, \delta))$

Then, it can be easily shown that H follows from P_1 and P_2, using the axioms $(\mathbf{Ax_2})$ and $(\mathbf{Ax_3})$.

5 Implementation and evaluation

To implement a full inference pipeline, one needs three components: (a) a syntactic parser that maps input sentences to CCG derivation trees, (b) a semantic parser that maps CCG derivation trees to SRs, and (c) a theorem prover that proves entailment relations between these SRs. In this study, we use manually constructed CCG trees as inputs and implement components (b) and (c).[10] For component (b), we use ccg2lambda[11] as a semantic parser and implement a set of templates corresponding to the lexical entries in Table 3. The system takes a CCG derivation tree as an input and outputs a logical formula as an SR. For component (c), we use the off-the-shelf theorem prover *Vampire*[12] and implement the set of axioms described in §4.

Suppose that the logical formulas corresponding to given premise sentences are $P_1, \ldots, P_n$ and that the logical formula corresponding to the hypothesis (conclusion) is H. Then, the system outputs *Yes* if $P_1 \wedge \cdots \wedge P_n \rightarrow H$ can be proved by a theorem prover, and outputs *No* if the negation of the hypothesis (i.e., $P_1 \wedge \cdots P_n \rightarrow \neg H$) can be proved. If both of them fail, it tries to construct a counter model; if a counter model is found, the system outputs *Unknown*. Since the main purpose of this implementation is to test the correctness of our semantic analysis and inference system, the system returns *error* if a counter model is not constructed with the size of an allowable model restricted.

We evaluate our system on the FraCaS test suite. The test suite is a collection of semantically complex inferences for various linguistic phenomena drawn from the literature on formal semantics and is categorized into nine sections. Out of the nine sections, we use ADJECTIVES (22 problems) and COMPARATIVES (31 problems). The distribution of gold answers is: (yes, no, unknown) = (9, 6, 7) for ADJECTIVES and (19, 9, 3) for COMPARATIVES. Table 6 lists some examples.

Table 5 gives the results of the evaluation. We compared our system with existing logic-based RTE systems. B&C (Bernardy and Chatzikyriakidis, 2017) is an RTE-system based on Grammatical Framework (Ranta, 2011) and uses the proof assistant Coq for theorem proving. The theorem proving part is not automated but manually checked. Nut (Bos, 2008) and MINE (Mineshima et al., 2015) use a CCG parser (C&C parser; Clark and Curran, 2007) and implement a theorem-prover for RTE based on FOL and higher-order logic, respectively. LP (Abzianidze, 2016) is a system, LangPro, that uses two CCG parsers (C&C parser and EasyCCG; (Lewis and Steedman, 2014)) and implements a tableau-based natural logic inference system. M&M (MacCartney and Manning, 2008)

[10] CCG parsers for English, such as C&C parser (Clark and Curran, 2007) based on CCGBank (Hockenmaier and Steedman, 2007), are widely used, but there is a gap between the outputs of these existing parsers and the syntactic structures we assume for the analysis of comparative constructions as described in §3. We leave a detailed comparison between those structures to another occasion. We also have to leave the task of combining our system with off-the-shelf CCG parsers for future research.

[11] https://github.com/mynlp/ccg2lambda

[12] https://github.com/vprover/vampire

Table 5: Accuracy on FraCaS test suite. '#All' shows the number of all problems and '#Single' the number of single-premise problems.

Section	#All	Ours	B&C	Nut	MINE	LP	M&M (#Single)
ADJECTIVES	22	**1.00**	.95	.32	.68	.73	.80* (15)
COMPARATIVES	31	**.94**	.56	.45	.48	-	.81* (16)

Table 6: Examples of entailment problems from the FraCaS test suite

FraCaS-198 (ADJECTIVES) Answer: No	
Premise 1	John is a former university student.
Hypothesis	John is a university student.
FraCaS-224 (COMPARATIVES) Answer: Yes	
Premise 1	The PC-6082 is as fast as the ITEL-XZ.
Premise 2	The ITEL-XZ is fast.
Hypothesis	The PC-6082 is fast.
FraCaS-229 (COMPARATIVES) Answer: No	
Premise 1	The PC-6082 is as fast as the ITEL-XZ.
Hypothesis	The PC-6082 is slower than the ITEL-XZ.
FraCaS-231 (COMPARATIVES) Answer: Unknown	
Premise 1	ITEL won more orders than APCOM did.
Hypothesis	APCOM won some orders.
FraCaS-235 (COMPARATIVES) Answer: Yes	
Premise 1	ITEL won more orders than APCOM.
Premise 2	APCOM won ten orders.
Hypothesis	ITEL won at least eleven orders.

uses an inference system for natural logic based on monotonicity calculus. M&M was only evaluated for a subset of the FraCaS test suite, considering single-premise inferences and excluding multiple-premise inferences. These four systems, Nut, MINE, LP, and M&M, are fully automated.

Although direct comparison is impossible due to differences in automation and the set of problems used for evaluation (single-premise or multiple-premise), our system achieved a considerable improvement in terms of accuracy. It should be noted that by using arithmetic implemented in Vampire our system correctly performed complex inferences from numeral expressions such as that in FraCaS-235 (see Table 6). Because we did not implement a syntactic parser and used gold CCG trees instead, the results show the upper bound of the logical capacity of our system. Note also that the five systems (B&C, MINE, LP, M&M, and ours) were developed in part to solve inference problems in FraCaS, where there is no separate test data for evaluation. Still, these problems are linguistically very challenging; from a linguistic perspective, the point of evaluation is to see *how* each system can solve a given inference problem. Overall, the results of evaluation suggest that a semantic parser based on degree semantics can, in combination with a theorem prover, achieve high accuracy for a range of complex inferences with adjectives and comparatives.

There are two problems in the COMPARATIVES section that our system did not solve: the inference from P to H_1 and the one from P to H_2, both having the gold answer *Yes*.

P: ITEL won more orders than the APCOM contract.
H_1: ITEL won the APCOM contract. (FraCaS-236)
H_2: ITEL won more than one order. (FraCaS-237)

To solve these inferences in a principled way, we will need to consider a more systematic way of handling comparative constructions that expects at least two patterns with missing verb phrases.

6 Conclusion

We proposed a CCG-based compositional semantics for gradable adjectives and comparatives using the A-not-A analysis studied in formal semantics. We implemented a system that maps CCG trees to suitable SRs and performs theorem proving for RTE. Our system achieved high accuracy on the sections for adjectives and comparatives in FraCaS.

In future work, we will further extend the empirical coverage of our system. In particular, we will cover deletion operations like Gapping in comparatives, as well as gradable expressions other than adjectives. Combining our system with a CCG parser is also left for future work.

Acknowledgement This work was supported by JSPS KAKENHI Grant Number JP18H03284.

References

Abzianidze, L. (2016). Natural solution to FraCaS entailment problems. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 64–74. Association for Computational Linguistics.

Bernardy, J.-P. and Chatzikyriakidis, S. (2017). A type-theoretical system for the FraCaS test suite: Grammatical framework meets Coq. In *Proceedings of the 12th International Conference on Computational Semantics (IWCS)*.

Bos, J. (2008). Wide-coverage semantic analysis with Boxer. In *Semantics in Text Processing. STEP 2008 Conference Proceedings*, pages 277–286.

Clark, S. and Curran, J. R. (2007). Wide-coverage efficient statistical parsing with CCG and log-linear models. *Computational Linguistics*, 33(4):493–552.

Cooper, R., Crouch, R., van Eijck, J., Fox, C., van Genabith, J., Jaspers, J., Kamp, H., Pinkal, M., Poesio, M., Pulman, S., et al. (1994). FraCaS–a framework for computational semantics. *Deliverable*, D6.

Cresswell, M. J. (1976). The semantics of degree. In *Montague Grammar*, pages 261–292. Elsevier.

Hackl, M. (2000). *Comparative Quantifiers*. PhD thesis, Massachusetts Institute of Technology.

Heim, I. (2000). Degree operators and scope. In *Semantics and Linguistic Theory*, volume 10, pages 40–64.

Hendriks, P. (1995). *Comparatives and Categorial Grammar*. PhD thesis, University of Groningen dissertation.

Hockenmaier, J. and Steedman, M. (2007). CCGbank: A corpus of CCG derivations and dependency structures extracted from the Penn Treebank. *Computational Linguistics*, 33(3):355–396.

Kennedy, C. (1997). *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison*. PhD thesis, University of California, Santa Cruz.

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and philosophy*, 4(1):1–45.

Klein, E. (1982). The interpretation of adjectival comparatives. *Journal of Linguistics*, 18(1):113–136.

Kubota, Y. and Levine, R. (2015). Against ellipsis: arguments for the direct licensing of 'noncanonical' coordinations. *Linguistics and Philosophy*, 38(6):521–576.

Larson, R. K. (1988). Scope and comparatives. *Linguistics and Philosophy*, 11(1):1–26.

Lassiter, D. (2017). *Graded Modality: Qualitative and Quantitative Perspectives*. Oxford University Press.

Lewis, M. and Steedman, M. (2014). A* CCG parsing with a supertag-factored model. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 990–1000. Association for Computational Linguistics.

MacCartney, B. and Manning, C. D. (2008). Modeling semantic containment and exclusion in natural language inference. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling)*, pages 521–528.

Martínez-Gómez, P., Mineshima, K., Miyao, Y., and Bekki, D. (2016). ccg2lambda: A Compositional Semantics System. In *Proceedings of ACL 2016 System Demonstrations*, pages 85–90.

Mineshima, K., Martínez-Gómez, P., Miyao, Y., and Bekki, D. (2015). Higher-order logical inference with compositional semantics. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2055–2061.

Narisawa, K., Watanabe, Y., Mizuno, J., Okazaki, N., and Inui, K. (2013). Is a 204 cm man tall or small? Acquisition of numerical common sense from the web. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 382–391. Association for Computational Linguistics.

Pulman, S. (2007). Formal and computational semantics: a case study. In *Proceedings of the Seventh International Workshop on Computational Semantics (IWCS)*, pages 181–196.

Ranta, A. (2011). *Grammatical Framework: Programming with Multilingual Grammars*. CSLI Publications.

Schwarzschild, R. (2008). The semantics of comparatives and other degree constructions. *Language and Linguistics Compass*, 2(2):308–331.

Seuren, P. A. (1973). The comparative. In *Generative Grammar in Europe*, pages 528–564. Springer.

Stechow, A. v. (1984). Comparing semantic theories of comparison. *Journal of Semantics*, 3(1-2):1–77.

Steedman, M. J. (2000). *The Syntactic Process*. The MIT Press.

van Rooij, R. (2008). Comparatives and quantifiers. *Empirical Issues in Syntax and Semantics*, 7:423–444.