The semantics and pragmatics of multi-head comparatives

Linmin Zhang, zhanglinmin@gmail.com, the version of December 23, 2023

Abstract

A multi-head comparative (e.g., *Fewer people own more land in this country than anywhere else in the world*, see also Chomsky 1981/1993, von Stechow 1984) contains two comparative expressions and one *than*-clause.

Semantically, I argue that there is a connection between a **cumulative-reading sentence** (e.g., *3% of the population owned 70% of the land*, see also Krifka 1999, Brasoveanu 2013) and a **multi-head comparative**. This connection is exactly parallel to that between a measurement sentence (e.g., *Mary is 6 feet tall*) and a comparative (e.g., *Sue is taller than Mary is*): the former means a **measurement**, while the latter means **comparison** between measurements. Therefore, a multi-head comparative is the comparative form of a cumulative-reading sentence, addressing comparisons along multiple dimensions.

I further propose that pragmatically, the interpertation of a multi-head comparative is based on one single underlying **degree question-under-discussion** (QUD) (e.g., for the above example: *how is wealth distribution more skewed in this country than anywhere else?*). Thus in a multi-head comparative, comparisons along multiple dimensions are not mutually independent, but address an interplay between increases or decreases along two dimensions, and it is this interplay that resolves the underlying degree QUD. I also discuss comparative correlatives and how multi-head comparatives shed light on the notion of informativeness.

Keywords: Comparatives, Multi-head comparatives, Cumulativity, Degree questions, QUD, Dynamic semantics, Maximality, Informativeness, Comparative correlatives

1 Introduction

A **multi-head comparative** contains two comparative expressions and one *than*-clause, as illustrated by the following examples (1)–(4), all from von Stechow (1984).

(1) More silly lectures have been given by more boring professors – than I would have expected. (p.42: (136a); originally from Chomsky 1981/1993: p.81, §2.4.4, (6ii))

(2) More dogs ate more rats than cats ate mice. (p.43: (140))

(3) Less land produces more corn than ever before. (p.46: (156))

(4) No airline saves you more money in more ways than Delta. (p.46: (157))

Although multi-head comparatives have been noticed for decades in formal syntax and semantics (see e.g., Chomsky 1981/1993, von Stechow 1984), a thorough investigation is still missing. This paper studies them from semantic and pragmatic perspectives, addressing how multi-head comparatives are built on other better-understood phenomena and how their interpretation sheds new light on the notion of informativeness.

According to von Stechow (1984), the semantics of multi-head comparatives should be a natural and ready extension of our existing understanding of comparatives:

> 'I don't think that these cases represent genuine semantic problems. They can essentially be treated with the methods we already have at our disposal. But it took me quite a while to realize this, because the examples are conceptually rather complicated and are neglected in the literature.'
>
> (von Stechow 1984, pp.41–42)

However, it has been questioned whether multi-head comparatives are, after all, interpretable and sensible (see e.g., Hendriks 1994, Hendriks and De Hoop 2001). Even if von Stechow (1984) denies that they present a real challenge, in his quotation above, he still characterizes them as 'conceptually rather complicated'.

Corpus search from *the Corpus of Contemporary American English* (https://www.english-corpora.org/coca/, Davies 2008-) suggests that there are indeed naturally occurring examples of multi-head comparatives:

(5) Henry Ford's assembly line in 1913 increased productivity. Fewer people made more stuff. (2002, NEWS: Atlanta Journal Constitution)

(6) Fewer and fewer people own more and more of the country's land. (2001, MAG: Christian Century)

(7) In the four years of the (Marshall) Plan, the Marshall agency spent $13.5 billion in 16 countries. Fewer people spent more money in that agency than ever before. (1998, TV: Cold War)

(8) As I moved through the years, I saw that I would make more money in more skilled positions. (2012, BLOG: pjmedia.com (http://pjmedia.com/tatler/2012/11/17/walmart-files-complaint-with-nlrb-over-union-harassment/))

Thus, descriptively, the current paper starts with a discussion on (i) what the intuitive interpretation of acceptable multi-head comparatives is and (ii) what unacceptable ones look like (Section 2). von Stechow (1984) points out that the comparative expressions in multi-head comparatives are restricted to plural or mass nouns (e.g., in (1), *more silly lectures* and *more boring professors* are plural nouns; in (3), *less land* and *more corn* are mass nouns) and suggests that this might be due to syntactic reasons. I also use empirical data to address whether there is indeed such a restriction.

Based on this description of empirical data, I show that there is a connection between multi-head comparatives and a well-understood phenomenon: **cumulative-reading sentences** (Section 3). I combine the studies of Krifka (1999), Brasoveanu (2013) and Zhang (2023) to propose the main ideas that lead to a formal analysis.

In a nutshell, semantically, I propose that multi-head comparatives are a combination of (i) comparatives and (ii) cumulative-reading sentences (see (9)):

(9) The connection between multi-head comparatives and other phenomena:

measurement sentence (expressing a single measurement)	$\rightsquigarrow$	**comparative sentence** (comparing single measurements)
Mary is d-tall		Sue is taller than Mary is ~~d-tall~~ (how tall Mary is)
cumulative-reading sentence (expressing a pair of measurements)	$\rightsquigarrow$	**multi-head comparative** (comparing pairs of measurements)
m-many cats ate n-many mice		More dogs ate more rats than ~~m-many~~ cats ate ~~n-many~~ mice (how many cats ate how many mice)

As a consequence, our existing understanding on cumulative-reading sentences can be carried over to multi-head comparatives, including (i) ambiguity between a cumulative and a distributive reading (see e.g., Brasoveanu 2013), (ii) the availability of a proportion-based cumulative reading in some context (see Krifka 1999), and (iii) the pragmatics in interpreting cumulative-reading sentences (see Krifka 1999, Zhang 2023).

Therefore, pragmatically, I propose that multi-head comparatives pattern with cumulative-reading sentences in addressing one single underlying degree QUD (Question under discussion, see Roberts 1996/2012). As illustrated in (10)–(13), there is an interplay between the two comparative expressions in a multi-head comparative, and together they provide information to resolve the underlying degree QUD.

(10) More silly lectures have been given by more boring professors – than I would have expected. (= (1))
⇝ addressing *how education is worse than I would have expected*

(11) More dogs ate more rats than cats ate mice. (= (2))
⇝ addressing *how dogs have more preys than cats do*

(12) Less land produces more corn than ever before. (= (3))
⇝ addressing *how corn productivity is better than before*

(13) No airline saves you more money in more ways than Delta. (= (4))
⇝ addressing *how Delta is more economical than other airlines*

A formal analysis that implements these ideas within dynamic semantics is presented in Section 4. Section 5 further compares the current work with existing studies (especially with regard to the notions of comparison and informativeness) and discusses how to to account for related phenomena (comparative correlatives and cross-linguistic data on multi-head comparatives). Section 6 concludes.

Basically, this paper supports von Stechow (1984)'s view that multi-head comparatives are not a challenge to our existing understanding of comparatives and 'can be treated with the methods we already have at our disposal'. The current paper achieves such an analysis by connecting multi-head comparatives with cumulative-reading sentences, demonstrating how natural language supports the expression of complex mathematical operations that range from one-dimensional to multi-dimensional comparison (see the upper vs. lower half of table (9)).

2 Empirical observations

This section presents a detailed empirical description of multi-head comparatives, focusing on three issues: (i) What is our intuitive interpretation of multi-head comparatives (Section 2.1)? (ii) What do unacceptable multi-head comparatives look like (Section 2.2)? (iii) In multi-head comparatives, are comparative expressions restricted to plural/mass nouns (Section 2.3)? This section is largely based on von Stechow (1984) and Hendriks (1994), Hendriks and De Hoop (2001)'s discussions. Further observations related to the connection between multi-head comparatives and cumulative-reading sentences are addressed in Section 3.

2.1 Intuitive interpretation of good multi-head comparatives

For a multi-head comparative like (2), von Stechow (1984) describes what it intuitively means and what it does not mean.

2.1.1 What a multi-head comparative means

As shown in (14), intuitively, the most easily available reading of sentence (2) means **two comparisons**, or in other words, point-wise two-dimensional comparison (see von Stechow 1984, p.43: (141)). Along the dimension of **agent cardinalities**, dogs that ate rats are compared with cats that ate mice (see (14a)). Along the dimension of **theme cardinalities**, rats eaten by dogs are compared with mice eaten by cats (see (14b)).

(14) von Stechow (1984): what a multi-head comparative intuitively means:
More dogs (agent) ate more rats (theme) than cats (agent) ate mice (theme). (= (2))

a. The comparison of **agent cardinalities**: |dogs| vs. |cats|
The number of dogs that ate rats > the number of cats that ate mice

b. The comparison of **theme cardinalities**: |rats| vs. |mice|
The number of rats eaten by dogs > the number of mice eaten by cats

For this reading sketched out in (14), there are three observations worthy of noting, which I list here and will come back to discuss more later.

Observation 1: the comparisons in (14) are point-wise, dimension by dimension. Specifically, there are two dimensions involved: (i) the cardinality of the agent (see (14a), |dogs| vs. |cats|); (ii) the cardinality of the theme (see (14b), |rats| vs. |mice|).

Observation 2: the dimensions involved are mutually independent. Thus, each comparison is also independent (see (14a) vs. (14b)).

Observation 3: there is mutual restriction between the agent and the theme, at the matrix-clause level and the *than*-clause level respectively. In interpreting (2)/(14), we do not count all the dogs and cats in our context and then compare the cardinalities of the absolutely maximal dog-sum and cat-sum. Rather, we count all the dogs that ate rats and all the cats that ate mice and then compare the cardinalities of the restricted dog-sum and cat-sum. Similarly, it is also for the restricted rat-sum (those eaten by dogs) and mouse-sum (those eaten by cats) that their cardinalities undergo comparison.

2.1.2 What a multi-head comparative does not mean

According to von Stechow (1984), (15) is not a good characterization for our intuitive interpretation of (2) (see von Stechow 1984, p.43: (142)). In contrast to (14), which expresses two mutually independent comparisons along two dimensions, (15) only expresses **one comparison**: the comparison between the cardinality of all dog-rat pairs involved in eating-events and that of all cat-mouse pairs involved in eating events.

(15) von Stechow (1984): what a multi-head comparative like (2) does not mean:
The number of pairs $\langle x, y\rangle$ > the number of pairs $\langle x', y'\rangle$
(where $\textsc{dog}(x), \textsc{rat}(y), \textsc{cat}(x'), \textsc{mouse}(y'), \textsc{eat}(x, y), \textsc{eat}(x', y')$)

This event-counting-based comparison reading in (15) is too weak, and this can be confirmed with a scenario like (16). Under the scenario (16), we intuitively judge sentence (2) false. However, the reading shown in (15) is actually true under this scenario, which is at odd with our intuition.

(16) Scenario: here are all the animal pairs involved in eating events:
$\langle \text{Dog}_1, \text{Rat}_1\rangle$, $\langle \text{Dog}_1, \text{Rat}_2\rangle$, $\langle \text{Dog}_1, \text{Rat}_3\rangle$, $\langle \text{Cat}_1, \text{Mouse}_1\rangle$, $\langle \text{Cat}_2, \text{Mouse}_2\rangle$
(Overall, 1 dog ate 3 rats, and 2 cats ate 2 mice between them. Thus here $|\langle x, y\rangle| = 3, |\langle x', y'\rangle| = 2$.)

2.2 Nonsensical multi-head comparatives

Hendriks (1994) (see also Hendriks and De Hoop 2001) points out that von Stechow (1984)'s view on multi-head comparatives would lead to over-generation, and that many multi-head comparatives predicted to be acceptable by von Stechow (1984) are in fact nonsensical and degraded, as illustrated by (17) and (18).

(17) *More doors (items) are higher (linear size) than windows (items) are wide (linear size). (Hendriks 1994: (5))

(18) *Fewer dogs (agent) ate more rats (theme) than cats (agent) ate mice (theme). (Hendriks 1994: (6))

As sketched out in (19) and (20), von Stechow (1984) predicts that (17) and (18) should be interpreted in a way similar to (14) and involve a pair of comparisons. However, according to Hendriks (1994), we intuitively feel that (17) and (18) sound weird, thus contradicting these predictions based on von Stechow (1984). The degradedness of (18) cannot be really due the opposite directions of the two inequalities (see (20)), given that data like (3) is intuitively natural and acceptable.

(19) The interpretation of (17) predicted by von Stechow (1984): two comparisons (see Hendriks 1994: (5'))

a. Along the dimension of item cardinalities:
The number of doors that are high > the number of windows that are wide

b. Along the dimension of linear size:
the height of high doors > the width of wide windows

(20) The interpretation of (18) predicted by von Stechow (1984): two comparisons (see Hendriks 1994: (6'))

a. Along the dimension of agent cardinalities:
The number of dogs that ate rats < the number of cats that ate mice

b. Along the dimension of theme cardinalities:
the number of rats eaten by dogs > the number of mice eaten by cats.

Why is von Stechow (1984)'s prediction not borne out? Among the three observations involved in the analysis in (14) (see Section 2.1.1), Hendriks (1994) considers Observation 2 problematic, arguing that the two comparisons involved in the

interpretation of a multi-head comparative cannot be mutually independent. Hendriks (1994)'s proposed analysis of (17) and (18) is shown in (21) and (22).

(21) Hendriks (1994)'s analysis of (17) (see Hendriks 1994: (5″)):
the number of doors x such that 'HEIGHT(x) > WIDTH(y)' >
the number of windows y such that 'HEIGHT(x) > WIDTH(y)'

(22) Hendriks (1994)'s analysis of (18) (see Hendriks 1994: (8)):
the number of dogs x such that '|rats eaten by x| > |mice eaten by y|' <
the number of cats y such that '|rats eaten by x| > |mice eaten by y|'

In (21) and (22), comparison along one dimension is embedded within comparison along another dimension. According to Hendriks (1994), for (21), it is the embedded comparison 'HEIGHT(x) > WIDTH(y)' that restricts the items x and y, and x and y are further compared along the dimension of cardinality. Hendriks (1994) claims that this restriction 'HEIGHT(x) > WIDTH(y)' makes x and y mutually dependent and results in infinite regress: the definition of x relies on the definition of y, and vice versa.[1]

Obviously, even if the degradedness of sentences like (17) and (18) challenges the analysis of von Stechow (1984) in (14), these data do not provide firm evidence in support of Hendriks (1994)'s analysis in (21) and (22). The degradedness of (17) and (18) might be due to other reasons (which I will explain later in Section 3.3). Besides, Hendriks (1994) cannot explain why multi-head comparatives like (1)–(4) are intuitively good and interpretable and do not suffer from the same problem as (17) and (18) do.

With regard to the degradedness of (17) and (18), Meier (2001) makes such a comment: 'The fact that (17a) (= (18) in the current paper) is somehow odd is just a reflex of the fact that it is hard to imagine a situation in which the comparison construction might be relevant (p.356).' In later sections I will connect multi-head comparatives with cumulative-reading sentences and adopt a similar view.

2.3 Are comparative expressions restricted to plural/mass nouns?

In addition to the nonsensical examples discussed above, some other sentences with multiple comparative expressions but one *than*-clause are intuitively unacceptable:

[1] In Sections 3 and 4, I will show that this kind of mutual restriction does not challenge semantic compositionality. Works like Brasoveanu (2013) have provided a solution (see also Charlow 2017). In addition to the literature on cumulative reading, Bumford (2017) also addresses mutual restriction and provides an analysis for the mutual definition of uniqueness in Haddock descriptions (see Haddock 1987).

(23) *More silly lectures have been given by more boring professors than I met yesterday. (Chomsky 1981/1993: p.81, §2.4.4, (6iii))

(24) *A greater man would be a better man than Otto. (von Stechow 1984: p.46, (158))

According to von Stechow (1984), the contrast between (23) and (1) (repeated here as (25)) suggests that each comparative expression needs to be associated with the *than*-clause, and this reconstruction of the *than*-clause for the first comparative expression (here *more silly lectures*) further motivates the two-comparison analysis (see (25)). (23) is different from (1)/(25) in that reconstructing the *than*-clause for the first comparative expression is impossible (see (26)).

(25) More silly lectures have been given by more boring professors – than I would have expected. (=(1))
⇝ More silly lectures (**than I would have expected that silly lectures would be given by boring professors**) have been given by more boring professors **than I would have expected** (**that silly lectures would be given by boring professors**).
⇝ Two comparisons: along the cardinalities of silly lectures and boring professors

(26) *More silly lectures (**than I met yesterday**) have been given by more boring professors **than I met yesterday**. Reconstructing the *than*-clause for (23)

It is along a similar reasoning that von Stechow (1984) explains why the comparative expressions in a multi-head comparative are restricted to plural or mass nouns. For a sentence like (24), the reconstruction gives rise to two possibilities in interpreting the *than*-phrase (see (27a) and (27b)). von Stechow (1984) claims that 'the low acceptability of (24) shows that our grammar doesn't work that way. It seems, then, that the restrictions for the reconstruction of a *than*-phrase are rather syntactic than semantic.'

(27) *A greater man (**than Otto**) would be a better man **than Otto**.
a. than Otto is a great man
b. than Otto is a good man

Thus, von Stechow (1984)'s view is that the reconstruction of a *than*-clause/phrase brings the requirement that the *than*-clause needs to serve as the comparison standard

for both comparative expressions, which further restricts the comparative expressions to be plural/mass nouns. von Stechow (1984) does not provide more details in explaining this restriction, but the problem is that his generalization about the restriction is empirically invalid.

As illustrated in (28), there can be good multi-head comparatives with a comparative expression that is not a plural/mass noun, here *faster*. This example (28) is interesting in two aspects. First, it is a counterexample to von Stechow (1984)'s generalization that in multi-head comparatives, comparative expressions are restricted to plural/mass nouns. Second, it also challenges the reconstruction approach proposed by von Stechow (1984). Under von Stechow (1984)'s reconstruction approach, (28) would be analyzed as (29), and *than before* serves as comparison standard for both *more good* and *faster*. Then if *than before* can be interpreted along two different scales/dimensions in (29), why cannot a similar mechanism work in the case of (27)?

(28) Nowadays, more goods are carried faster (than before). (Hendriks and De Hoop 2001: p.10, (13))

(29) Nowadays, more good (**than before**) are carried faster **than before**.

2.4 Interim summary

To sum up the above discussion, for a good multi-head comparative, its most easily available reading involves multiple comparisons, or in other words, point-wise multi-dimensional comparison. However, this analysis leads to over-generation, and not all sentences containing multiple comparative expressions are intuitively interpretable.

Along the discussion, it becomes evident that degraded examples of multi-head comparatives are not necessarily due to syntactic constraints (see (18) vs. (2), which have parallel syntax) or the opposite directions of inequalities (see (18) vs. (3)). Besides, comparative expressions in a multi-head comparative are not always restricted to plural/mass nouns (see (28)).

Below I start with Observation 3 in 2.1.1: the mutual restriction at the matrix-clause level and the *than*-clause level respectively. I show that something similar can be observed on a well-understood construction, cumulative-reading sentences, and propose the idea that multi-head comparatives are actually the comparative form built on cumulative-reading sentences.

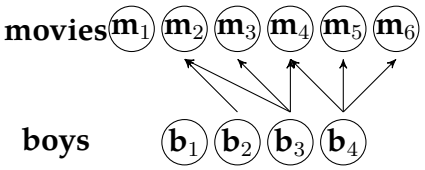

Figure 1: The genuine **cumulative** reading of (30) is **true** in this context.

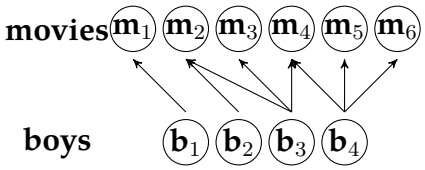

Figure 2: The genuine **cumulative** reading of (30) is **false** in this context.

3 Inspiration from cumulative-reading sentences

Here I first present Brasoveanu (2013)'s analysis of cumulative-reading sentences (Section 3.1). Then I follow Krifka (1999) and Zhang (2023)'s discussion to show the pragmatics involved in the interpretation of cumulative-reading sentences (Section 3.2). Based on this, I informally account for the data discussed in Section 2 (Section 3.3).

3.1 Brasoveanu (2013)'s analysis of cumulative-reading sentences

A sentence like (30) contains two modified numerals (see the underlined parts) and has both a **distributive** and a **cumulative** reading (see (30a) and (30b)). Here I focus on the cumulative reading. (The reading of (30a) will be addressed in Section 4.6).

(30) Exactly three boys saw exactly five movies. (see e.g., Brasoveanu 2013)

a. **Distributive**: There are exactly 3 boys s.t. each of them saw exactly 5 movies. (In total, there are $3 \times 5 = 15$ movie-seeing events. The number of distinct movies involved is between 5 and 15).
$\rightsquigarrow$ 'exactly 3 boys' takes scope over 'exactly 5 movies'

b. **Cumulative**: The total number of boys who saw any movies is 3, and the total number of movies seen by any boys is 5. (In total, the number of movies involved is 5, and there are between 5 and 15 movie-seeing events).
$\rightsquigarrow$ **No scope taking between the two modified numerals**

Krifka (1999), Brasoveanu (2013), Charlow (2017) emphasize that the derivation of the genuine cumulative reading (30b) is necessarily built on **mutual restriction** between the two modified numerals, and that no scope-taking is involved.

Suppose that there is no mutual restriction. As illustrated in (31), one modified numeral (here *exactly three boys*) takes scope over the other (here *exactly five movies*), so

that only the latter is involved in restricting and defining the former, but not vice versa. The reading derived (dubbed as the pseudo-cumulative reading) is true under the scenario in Fig. 2, as shown in (31).

(31) **Unattested pseudo-cumulative reading** of (30): The maximal boy-sum such that 'they saw in total five movies between them' has the cardinality 3.
⇝ True under the scenario in Fig. 2: there are two boy-sum witnesses, $\mathbf{b}_2 \oplus \mathbf{b}_3 \oplus \mathbf{b}_4$ and $\mathbf{b}_1 \oplus \mathbf{b}_2 \oplus \mathbf{b}_4$. There is no larger boy-sum satisfying the restriction that 'they saw in total five movies between them'.

Intuitively, sentence (30) is false under the scenario in Fig. 2, indicating that the pseudo-cumulative reading in (31) is too weak and unattested. Thus the genuine cumulative reading of (30) must be stronger and involve mutual restriction.

The genuine cumulative reading (see (30b)) is true under the scenario in Fig. 1. For this reading, *exactly three boys* denotes and counts the totality of **boys who saw any movies**, which is 3 (cf. the cardinality of all boys in the context is 4!), and *exactly five movies* denotes and counts the totality of **movies seen by any boys**, which is 5 (cf. the cardinality of all movies in the context is 6!).

Brasoveanu (2013) adopts dynamic semantics to implement the above ideas and develop a formal analysis for the cumulative reading of (30). Essentially, modified numerals make semantic contribution in several layers, as sketched out in (32):

(32) Exactly three boys saw exactly five movies. **Cumulative reading** of (30)

$$\underbrace{\sigma x \sigma y[\textsc{boy}(x) \wedge \textsc{movie}(y) \wedge \textsc{saw}(x,y)]}_{\text{the mereologically maximal } x \text{ and } y} \wedge \underbrace{|y| = 5 \wedge |x| = 3}_{\text{cardinality tests}}$$

First, each modified numeral introduces a potentially plural discourse referent (dref). Then restrictions are added onto drefs: here $\textsc{boy}(x)$, $\textsc{movie}(y)$, and $\textsc{saw}(x,y)$.[2]

Second, at the sentence level, when all the drefs are introduced and restrictions are added, modified numerals contribute **mereology-based maximality operators**. These maximality operators are applied at the sentence level simultaneously, picking out the maximal plural individuals satisfying the restrictions $\textsc{boy}(x)$, $\textsc{movie}(y)$, and $\textsc{saw}(x,y)$.

Finally, the modified numerals contribute **cardinality tests**, checking whether the cardinality of the relativized maximal boy-sum (i.e., those who saw movies) is 3 and whether the cardinality of the relativized maximal movie-sum is 5.

[2]Cumulative closure is assumed for lexical relations when needed.

Evidently, there is great similarity between cumulative-reading sentences and multi-head comparatives. Our intuitive interpretation for both of them involves mutual restriction and relativized maximization. Mutual restriction paves ground for the selection / definition of relativized maximal plural individuals. Then cumulative-reading sentences **address the cardinality of each relativized maximal plural individual** (see (33a)), while multi-head comparatives **compare their cardinalities** (see (33b)).

(33) Measurement and comparison along two dimensions: the cardinalities of agent and theme

a. Exactly three cats ate exactly five mice.
   ⇝ how many cats ate how many mice?

b. More dogs ate more rats $\underbrace{\underbrace{\text{than cats ate mice}}_{\text{how many cats ate how many mice}}}_{\text{how many dogs ate how many rats}}$. (= (2)/(14))

Thus the connection between cumulative-reading sentence (33a) vs. multi-head comparative (33b) is parallel with that between measurement sentence (34a) vs. comparative (34b). For usual measurement and comparatives sentences, measurement and comparison are conducted along one single dimension (e.g., height in (34)). For cumulative-reading sentences and multi-head comparatives, measurement and comparison are conducted simultaneously along multiple dimensions (e.g., the cardinalities of agent and theme in (33)). Multi-head comparatives are actually the comparative form of cumulative-reading sentences.

(34) Measurement and comparison along one dimension: height

a. Mary is 6 feel tall. ⇝ How tall is Mary?

b. Sue is taller $\underbrace{\underbrace{\text{than Mary is } \sout{\text{tall}}}_{\text{how tall Mary is}}}_{\text{how tall Sue is}}$.

Analyzing multi-head comparatives as the comparative form of cumulative-reading sentences (see (33b)) captures immediately Observations 1 and 3 in von Stechow (1984)'s analysis (see (14) in Section 2.1.1): Comparison is conducted along multiple dimensions, and there is mutual restriction between the agent and the theme at both the matrix-clause and the *than*-clause level (see (33b)). Below I discuss whether the multiple dimensions involved are mutually independent or not (Observation 2).

3.2 Krifka (1999) and Zhang (2023) on cumulative-reading sentences

Krifka (1999) discusses data like (35), another kind of cumulative-reading sentences, and points out that the **simultaneous mereology-based maximization** strategy (that works for the analysis of cumulative-reading sentences like (30)) cannot work for (35):

(35) In Guatemala, (at most) 3% of the population own (at least) 70% of the land.

Krifka (1999) points out that the simultaneous mereology-based maximization

> 'would lead us to select the alternative *In Guatemala, 100 percent of the population own 100 percent of the land*, which clearly is not the most informative one among the alternatives – as a matter of fact, it is pretty uninformative.'
>
> ...
>
> 'What is peculiar with sentences like (35) is that they want to give information about the bias of a statistical distribution. One conventionalized way of expressing particularly biased distributions is to select a small set among one dimension that is related to a large set of the other dimension.'
>
> (Krifka 1999: §3.1)

Krifka (1999)'s discussion suggests that the use of modified numerals in a cumulative-reading sentence addresses an underlying Question under discussion (QUD, see Roberts 1996/2012). The utterance of a sentence like (35) aims to 'express particularly biased distribution', and in expressing **how biased this distribution is** – a degree QUD, the utterance of (*at most*) *3% of the population* and (*at least*) *70% of the land* is more informative than its alternatives. Thus, in interpreting modified numerals and cumulative-reading sentences, maximization needs to be informativeness-based, not necessarily mereology-based.

Based on this idea, Zhang (2023) provides a unified analysis of (30) and (35), with the use of **informativeness-based maximality operators** (see Section 4). The interpretation of a cumulative-reading sentence assumes a **single** underlying degree QUD, which is addressed by the **interplay among multiple modified numerals**.

A cumulative-reading sentence like (30) is most naturally interpreted as addressing *how high the overall film consumption among boys is*. As illustrated by Fig. 3, the maximally informative true answer to this degree question corresponds to the cardinalities of the mereologically maximal drefs, i.e., the right-uppermost dot in Fig. 3.

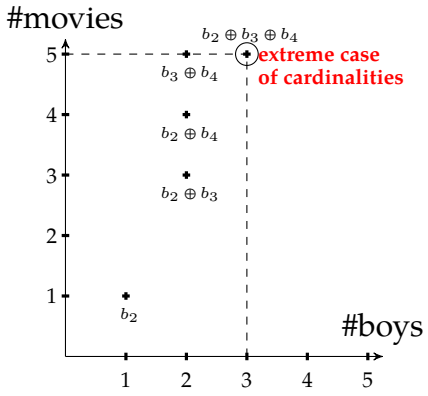

Figure 3: (30): Exactly three boys saw exactly five movies.
The plotting of some boy-sums and the total amount of movies they saw (in the context of Fig. 1) is shown here.
For the degree QUD *how high the overall film consumption among boys is*, the most informative answer is represented by the right-uppermost dot, i.e., the one corresponding to the boy-sum $b_2 \oplus b_3 \oplus b_4$ and the 5 movies they saw between them.

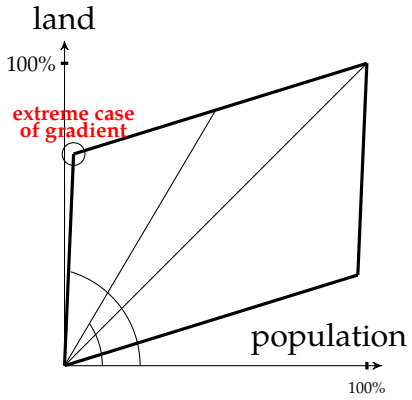

Figure 4: (35): In Guatemala, (at most) 3% of the population own (at least) 70% of the land.
The plotting of the percentages of land-owning populations and their total owned land is shown as a parallelogram-like area.
For the degree QUD *how skewed wealth distribution in Guatemala is*, the most informative answer is represented by the left-uppermost corner, i.e., at most 3% of the population own at least 70% of the land.

On the other hand, a cumulative-reading sentence like (35) is most naturally interpreted as addressing *how skewed wealth distribution in Guatemala is*. As illustrated by Fig. 4, the maximally informative true answer to this degree question corresponds to the case with the maximal ratio between the amount of land and the population of its owner, i.e., the left-uppermost corner in Fig. 4.

Evidently, the interpretation of these multi-head comparatives illustrates how human language uses the overall interplay among multi-dimensional measurements to address a single degree QUD. Mereology-based maximality, which is involved in Brasoveanu (2013)'s analysis, is a special case of this degree-QUD-based maximality. The simultaneity in applying two maximality operators (see (32)) is actually not a stipulation, but rather a natural consequence of applying informativeness-based maximization: in Fig. 3, the right-uppermost point represents the case where measurements along both dimensions are maximal.

The overall interplay among multi-dimensional measurements in addressing an

underlying degree QUD also explains the monotonicity of modified numerals involved in a cumulative-reading sentence, i.e., whether the increase or decrease of numerals leads to higher informativeness.

For sentences like (30), increase along both dimensions leads to higher informativeness. In other words, the two numerals in (30) contribute to the informativeness of the whole sentence in a parallel way, and thus they share the same monotonicity (see the right-uppermost corner in Fig. 3).

For sentences like (35), increase along one dimension and decrease along another dimension together lead to higher informativeness. In other words, the two numerals in (35) contribute to the informativeness of the whole sentence in opposite ways, and thus they can have different monotonicity: *at most 3%* is downward-entailing, while *at least 70%* is upward-entailing (see also the left-uppermost corner in Fig. 4).

3.3 From cumulative-reading sentences to multi-head comparatives

The connection between cumulative-reading sentences and multi-head comparatives captures von Stechow (1984)'s Observation 1 (multi-dimensional comparison) and Observation 3 (mutual restriction at each level of the matrix and embedded *than* clauses). The degree-QUD-based pragmatics involved in interpreting cumulative-reading sentences accounts for Hendriks (1994)'s observation: the two comparisons in a multi-head comparative are not mutually independent.

Just like cumulative-reading sentences, a multi-head comparative uses its multi-dimensional comparison to address one single underlying degree QUD. As illustrated in (36)–(39) and Fig. 5-6, the interplay between the increase or decrease along multiple dimensions resolves the underlying degree QUD of these multi-head comparatives. Usually, the use of two **positive** comparative expressions (see e.g., (36) and Fig. 5) addresses an underlying degree QUD with an overall increase, indicating a right-upward movement on the plotting. The use of two comparative expressions with **opposite polarities** (see e.g., (38) and Fig. 6) addresses an underlying degree QUD with an increase of ratio, indicating a left-upward movement on the plotting.

(36) More silly lectures have been given by more boring professors – than I would have expected. (= (1)/(10))
⇝ QUD: *how education is worse than I would have expected*
Increase along both the quantity of silly lectures and boring professors leads to

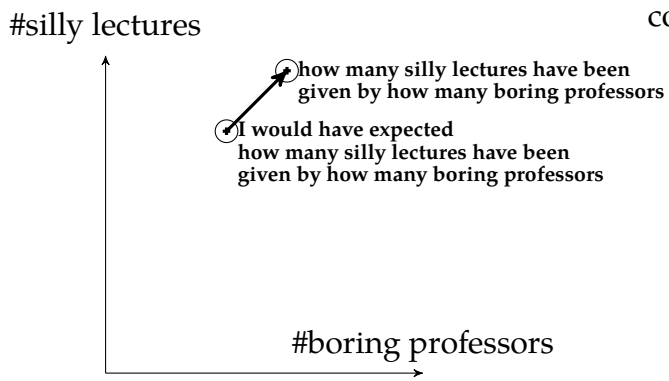

Figure 5: (1): More silly lectures have been given by more boring professors – than I would have expected.
To resolve the underlying QUD *how education is worse than I would have expected*, (1) addresses increases along multiple dimensions, i.e., an overall increase (see also (36)).

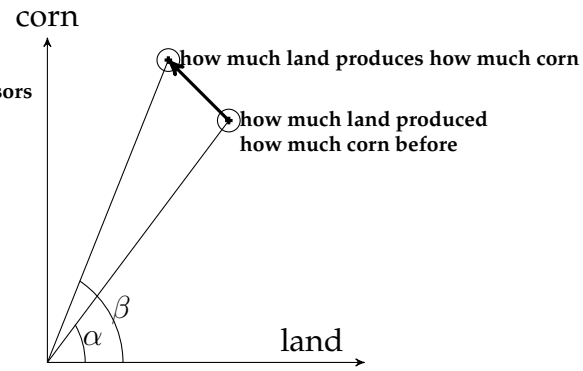

Figure 6: (3): Less land produces more corn than ever before.
To resolve the underlying QUD *how corn production is better than before*, (3) addresses the interplay between the decrease along the dimension of land quantity and the increase along the dimension of corn production, i.e., an increase of ratio ($\beta > \alpha$, see also (38)).

higher informativeness in resolving the underlying QUD.

(37) More dogs ate more rats than cats ate mice. (= (2)/(11))
⇝ QUD: *how dogs have more preys than cats do*
Increase along both the quantity of dogs and rats leads to higher informativeness in resolving the underlying QUD.

(38) Less land produces more corn than ever before. (= (3)/(12))
⇝ QUD: *how corn productivity is better than before*
Increase along the quantity of corn and decrease along the quantity of land together lead to higher informativeness in resolving the underlying QUD.

(39) No airline saves you more money in more ways than Delta. (= (4)/(13))
⇝ QUD: *how Delta is more economical than other airlines*
Increase along both the quantity of money and ways of saving leads to higher informativeness in resolving the underlying QUD.

This degree-QUD-based pragmatics in interpreting multi-head comparatives is in line with Meier (2001)'s comments: for degraded multi-head comparatives, 'it's hard to

imagine a situation in which the comparison construction might be relevant (p.356).' In other words, for degraded examples like (17) and (18), a natural underlying degree QUD is missing. However, similar examples should be felicitous when a contextually salient QUD is available, as illustrated in (40)–(42).

(40) Aldi's pork selection was a little more varied with better value than what I saw here. (cf. (17))
(A naturally occurring example: https://www.insider.com/aldi-vs-lidl-review-differences-which-better-photos-2021-5)
QUD: *how Aldi's pork selection was better*

(41) Hydrogen-powered cars also provide slightly more driving range with better energy density than batteries and fuel. (cf. (17))
(A naturally occurring example: https://www.whichcar.com.au/car-advice/hydrogen-cars-v-electric-cars-australia)
QUD: *how hydrogen-powered cars use energy more efficiently*

(42) Context: A zoo keeper is talking about the cost of feeding animals. Lions eat between 5 to 7 kg of meat a day, while a chimpanzee typically eats fruit, and the amount is between 1 to 4 kg a day. Feeding lions is more costly.[3]
Fewer lions eat more meat than chimpanzees eat fruit. (cf. (18))
QUD: *how much more costly feeding lions is than feeding chimpanzees in a zoo*

Examples (40) and (41) are also similar to (28) (repeated here as (43)) in not meeting the 'plural/mass noun' restriction for comparative expressions (cf. von Stechow 1984). These sentences are felicitous due to the availability of a salient degree QUD.

(43) Nowadays, more goods are carried faster (than before). (= (28))
⇝ QUD: *how transportation is more efficient than before*

Other degraded examples addressed by von Stechow (1984) (which attributes their degradedness to syntactic reasons) can also be accounted for pragmatically. A multi-head comparative is actually similar to usual comparatives in being able to have a large ellipsis part in their *than*-clause (see e.g., (43)). Ample examples have shown that

[3] I thank Takeo Kurafuji for discussing the example (18) with me and providing this scenario in (42). In Section 4.6, I will further discuss the non-cumulative reading of sentences like (42) (and (40)-(43)). Here these examples show that our intuitive acceptance of a multi-head comparative depends on the availability of a coherent degree QUD.

multiple gradable adjectives and their dimensions can be recovered from this kind of ellipsis within a *than*-clause (see e.g., (40)–(42)). Thus, for degraded examples like (23) and (24) (repeated here as (44) and (45)), the seeming difficulty of reconstructing a *than*-clause for each comparative expression should be considered fundamentally due to the lack of a salient underlying degree QUD. In (44), the matrix clause addresses *how many silly lectures have been given by how many boring professors* (or *how bad education is*), while the *than*-clause does not. In (45), the matrix clause addresses a correlation along two dimensions, while the *than*-expression part does not. It is this asymmetry of the underlying QUD between the matrix clause and the *than*-clause that leads to the degradedness of (44) and (45). When the matrix and the *than*-clause are parallel in addressing an underlying degree QUD and together address a change that involves multiple dimensions (e.g., increase of production rate in (3)), the recovery of content for the ellipsis in a *than*-clause/phrase does not lead to problems (see e.g., (40) and (46)).

(44) *More silly lectures have been given by more boring professors than I met yesterday. (= (23), see also (26))

(45) *A greater man would be a better man than Otto. (= (24), see also (27))

(46) She is a greater person with more determination than Otto.

After presenting the formal analysis for multi-head comparatives in Section 4, I will make further comparisons between the current proposal and existing works in Section 5.

4 Formal analysis of multi-head comparatives

This section provides a formal analysis for multi-head comparatives like (2). As sketched out in (47), the semantics of multi-head comparatives combines the semantics of comparatives and cumulative-reading sentences:

(47) The comparative form of a cumulative-reading sentence:
More dogs ate more rats than $\underbrace{\underbrace{\textit{m}\text{-}\sout{\text{many}}\ \text{cats ate}\ \textit{n}\text{-}\sout{\text{many}}\ \text{mice}}_{\textbf{\textit{than}-clause}\text{: how many cats ate how many mice}}}_{\textbf{matrix clause}\text{: how many dogs ate how many rats}}$. (= (2))
QUD: *how dogs have more preys than cats do* (or *how dogs are more successful*)

For the semantics of comparatives (see Section 4.1), I adopt the insights from Zhang

and Ling (2021): (i) the *than*-clause is the most informative true answer to a degree question (see also Fleisher 2018, 2020), and (ii) the matrix clause indicates an increase, based on the scalar value(s) provided by the *than*-clause. For the semantics of cumulative-reading sentences (see Section 4.2), I follow Brasoveanu (2013)'s analysis within Dynamic Predicate Logic (DPL, Groenendijk and Stokhof 1991).[4] Section 4.3 combines the semantics of comparatives and cumulative-reading sentences to derive the meaning of a multi-head comparative like (2)/(47). Section 4.4 zooms into the notion of maximal informativeness, presenting a more generalized, degree-QUD-based maximality operator. Based on this, Sections 4.5 and 4.6 further address two extensions: multi-head comparatives with a proportion reading or a non-cumulative reading.

4.1 The semantics of comparatives

According to Zhang and Ling (2021), as illustrated in (48), in an English comparative, (i) the *than*-clause is the most informative true answer to a corresponding degree question (here *how tall Mary is*), (ii) English comparative morpheme *-er/more* denotes an **increase** on a base value, and (iii) a comparative sentence denotes an increase on the base of the meaning of the *than*-clause.

(48) Sue is tall $\underbrace{\underbrace{\text{er}}_{\text{an increase}} \underbrace{\text{than Mary is } \sout{d\text{-tall}}}_{\textbf{\textit{than}}\textbf{-clause: how tall Mary is – the base for an increase}}}_{\textbf{matrix clause: how tall Sue is – an increase on the base of the } \textit{than}\text{-clause}}$.

Analyzing comparative morpheme *-er/more* as an increase is motivated by the parallelism between *-er/more* and additive particles like *(an)other* (see also Greenberg 2010, Thomas 2010, 2011 for relevant data and discussion):

(49) *-er/more* vs. *(an)other*: the increase and its base are in different sentences:

a. I ate $\underbrace{\text{an apple}}_{\text{the base}}$. Then I ate $\underbrace{\textbf{another}}_{\text{an increase}}$.

b. I ate $\underbrace{\text{some chocolate}}_{\text{the base}}$. Then I ate $\underbrace{\text{(a bit) } \textbf{more} \text{ (chocalate)}}_{\text{an increase}}$

c. Mary is tall. Sue is tall $\underbrace{\textbf{er}}_{\text{an increase based on the height of Mary}}$.

[4] Charlow (2014, 2017), Bumford (2017) use monads to develop a more compositional formalism that captures the spirit of Brasoveanu (2013)'s analysis. Since this issue of compositionality is orthogonal to the main goal of the paper, I follow Brasoveanu (2013)'s formalism within DPL, with some slight notation differences.

(50) *-er/more* vs. *(an)other*: the increase and its base are in the same sentence:

a. $\underbrace{\text{A girl, Mary}}_{\text{the base and its restriction}}$, met $\underbrace{\textbf{another girl, Sue}}_{\text{an increase and its restriction}}$.

b. Sue is $\underbrace{\text{(2 inches)}}_{\text{the restriction of the increase}}$ tall $\underbrace{\textbf{er}}_{\text{an increase}}$ $\underbrace{\text{than Mary}}_{\text{the base}}$.

Thus, Zhang and Ling (2021) defines comparative morpheme *-er/more* as an unspecified **positive scalar value** (i.e., an **increase**), with a **requirement of additivity**, i.e., there is a contextually salient item serving as the base for this increase. Specifically, within an interval-theoretic framework (see (51a)), ⟦-er/more⟧ denotes the most general positive interval, $(0, +\infty)$.[5] In this paper (except for Section 4.6), I use a simplified degree-based formalism (see (51b)): ⟦-er/more⟧ denotes an unspecified positive degree.

(51) a. $[\![\text{-er/more}]\!]_{\langle dt\rangle} \stackrel{\text{def}}{=} (0, +\infty)$ an **interval** (see Zhang and Ling 2021)

b. $[\![\text{-er/more}]\!]_{d} \stackrel{\text{def}}{=} d_0$ such that $d_0 \in (0, +\infty)$ an unspecified positive **degree**

The derivation of the semantics of a comparative is shown in (52).

(52) $\underbrace{\text{Sue is tall}\underbrace{\text{er}}_{d_0}\ \underbrace{\text{than Mary is } \sout{\text{tall}}}_{\text{HEIGHT(Mary)}}}_{\text{HEIGHT(Sue)} \geq d_0 + \text{HEIGHT(Mary)}}$ (= (48))

a. **The semantics of *tall***: $[\![\text{tall}]\!]_{\langle d,et\rangle} \stackrel{\text{def}}{=} \lambda\delta_d.\lambda x_e.\textsc{height}_{\langle e,d\rangle}(x) \geq \delta$
(⟦tall⟧ relates a degree δ and an individual x: the height of x reaches δ)

b. **The semantics of the *than*-clause**:
$\textsc{max}[\lambda d.$ Mary is d ~~tall~~$] = \textsc{max}[\lambda d.\textsc{height}(\text{Mary}) \geq d]$
= the maximal degree d that the height of Mary reaches = $\textsc{height}(\text{Mary})$
(here $\textsc{max} \stackrel{\text{def}}{=} \lambda D_{\langle dt\rangle}.\iota d[d \in D \wedge \forall d'[d' \in D \rightarrow d' \leq d]]$)

c. **The semantics of the matrix clause**: $\textsc{height}(\text{Sue}) \geq d_0 + \textsc{height}(\text{Mary})$
i.e., the height of Sue reaches 'an unspecified increase based on the height of Mary'.[6]

[5] An interval is a convex set of degrees (A totally ordered set S is convex if for any two items $a, b \in S$ (suppose $a < b$), then $\forall x[a \leq x \leq b \rightarrow x \in S]$.). Given that the type of degrees is d, the type of intervals is $\langle dt\rangle$. $(0, +\infty)$ is the shorthand of $\{d \mid 0 < d < +\infty\}$ (see Zhang and Ling 2021 for more details).

[6] We can assume a silent operator Compare: it computes the degree input for the gradable adjective *tall* in the matrix clause from (i) the base of the increase (i.e., the comparative standard – the meaning of the *than*-clause) and (ii) the increase: $\textsc{Compare}_{\langle d,\langle d,d\rangle\rangle} \stackrel{\text{def}}{=} \lambda d_{\text{INCREASE}}.\lambda d_{\text{BASE}}.d_{\text{INCREASE}} + d_{\text{BASE}}$. But actually, this assumption of Compare is conceptually not needed, because it is already concealed in the additivity requirement of comparative morpheme *-er/more*.

4.2 The semantics of cumulative-reading sentences

Brasoveanu (2013) provides a formal analysis of cumulative-reading sentences within Dynamic Predicate Logic (DPL, Groenendijk and Stokhof 1991). The gist is that modified numerals make semantic contribution in several layers: (i) introducing drefs; (ii) imposing mereology-based maximality operators; and (iii) checking cardinalities.

Within dynamic semantics, meaning derivation is considered a series of updates from an information state to another. Given the distributivity of DPL (see (53b), Groenendijk and Stokhof 1990), an update is a relation between assignment functions (i.e., it takes an assignment function as input and returns a set of assignment functions).

(53) DPL (see Groenendijk and Stokhof 1990, 1991):

a. **Information state** i: a set of assignment functions g Type: $\langle st\rangle$

b. **Update** r: from an information state to another Type: $\langle st, st\rangle$
Given that for every information state i, $r(i) = \bigcup_{g\in i} r(\{g\})$, an update can be considered a relation between assignment functions (of type $\langle s, st\rangle$).

c. **Truth**: an update is true if it does not end with an empty set.

The semantic derivation of the cumulative-reading is illustrated in (54):

(54) $[\![$Exactly three^u boys saw exactly five$^\nu$ movies$]\!]$ (=(30)/(32))

$\Leftrightarrow$ $\underbrace{\mathbf{3}_u[\mathbf{5}_\nu}_{\text{cardinality tests}}$ $[$ $\underbrace{\mathbf{M}_{u,\nu}}_{\text{maximality}}$ $\underbrace{[\![\text{some}^u \text{ boys saw some}^\nu \text{ movies}]\!]}_{\text{dref introduction}}]]$

a. **Introducing drefs**: $[\![$some^u boys saw some$^\nu$ movies$]\!]$
$\Leftrightarrow \lambda g.\left\{ g^{\substack{\nu\mapsto y\\ u\mapsto x}} \middle|\, \textsc{movie}(y), \textsc{boy}(x), \textsc{saw}(x,y)\right\}$

b. **Simultaneously applying mereology-based maximality operators**:
$\mathbf{M}_{u,\nu} \overset{\text{def}}{=} \lambda m_{\langle s,st\rangle}.\lambda g_s.\{h \in m(g) \mid \neg\exists h' \in m(g).h(u) \sqsubset h'(u) \vee h(\nu) \sqsubset h'(\nu)\}$
Thus $\mathbf{M}_{u,\nu}[\![$some^u boys saw some$^\nu$ movies$]\!]$
$\Leftrightarrow \mathbf{M}_{u,\nu}[\lambda g.\left\{ g^{\substack{\nu\mapsto y\\ u\mapsto x}} \middle|\, \textsc{movie}(y), \textsc{boy}(x), \textsc{saw}(x,y)\right\}]$
$\Leftrightarrow \lambda g.\left\{ g^{\substack{\nu\mapsto y\\ u\mapsto x}} \middle|\, \begin{array}{l} y = \Sigma y[\textsc{movie}(y) \wedge \textsc{saw}(x,y)] \\ x = \Sigma x[\textsc{boy}(x) \wedge \textsc{saw}(x,y)] \end{array} \right\}$

c. **Checking cardinalities**:
$\mathbf{3}_u \overset{\text{def}}{=} \lambda m_{\langle s,st\rangle}.\lambda g_s.\begin{cases} m(g) & \text{if } |g(u)| = 3 \\ \varnothing & \text{otherwise} \end{cases}$

$$\mathbf{5}_\nu \overset{\text{def}}{=} \lambda m_{\langle s,st\rangle}.\lambda g_s.\begin{cases} m(g) & \text{if } |g(\nu)| = 5 \\ \varnothing & \text{otherwise} \end{cases}$$

Thus $\mathbf{3}_u[\mathbf{5}_\nu[\mathbf{M}_{u,\nu}[\![\text{some}^u \text{ boys saw some}^\nu \text{ movies}]\!]]]$

$$\Leftrightarrow \lambda g.\left\{ g^{\substack{\nu\mapsto y \\ u\mapsto x}} \middle|\; \begin{array}{l} y = \Sigma y[\textsc{movie}(y) \wedge \textsc{saw}(x,y)] \\ x = \Sigma x[\textsc{boy}(x) \wedge \textsc{saw}(x,y)] \end{array} \right\}, \text{ if } |x| = 3 \wedge |y| = 5$$

In (54), modified numerals first work like existential quantifiers and introduce drefs, and these drefs get restrictions like $\textsc{movie}(y), \textsc{boy}(x), \textsc{saw}(x,y)$ (see (54a)). Then mereology-based maximality operator $\mathbf{M}_{u,\nu}$ picks out the maximal boy-sum and movie-sum that satisfy all the relevant restrictions (see (54b)). Finally, $\mathbf{3}_u$ and $\mathbf{5}_\nu$ check the cardinalities of these maximal boy-sum and movie-sum. Eventually, the cumulative reading of (30)/(32) is an update such that it is true if the cardinality of all boys who saw movies is 3 and the cardinality of all movies seen by boys is 5.

4.3 The semantics of multi-head comparatives

As illustrated in (55), the semantics of a multi-head comparative like (2) is directly built upon the meaning of cumulative-reading sentences (see (55a)–(55c)) and comparatives (see (55c)).

(55) $\underbrace{\text{More}^u}_{\text{an increase: } d_1}$ dogs ate $\underbrace{\text{more}^\nu}_{\text{an increase: } d_2}$ rats $\underbrace{\text{than } \sout{m\text{-many}}^{u'} \text{ cats ate } \sout{n\text{-many}}^{\nu'} \text{ mice}}_{\text{the cardinalities of relevant cats and mice: } |g(u')|,|g(\nu')|}$ (= (2))

$\Leftrightarrow$

$$\underbrace{\mathbf{more}_{u,u'}}_{\text{comparing cardinalities}}[\underbrace{\mathbf{more}_{\nu,\nu'}}_{}[\underbrace{\mathbf{M}_{u,\nu,u',\nu'}}_{\text{maximality}}\underbrace{[\![\text{some}^u \text{ dogs ate some}^\nu \text{ rats, some}^{u'} \text{ cats ate some}^{\nu'} \text{ mice}]\!]}_{\text{dref introduction}}]]$$

a. **Introducing drefs**:

$[\![\text{some}^u \text{ dogs ate some}^\nu \text{ rats, some}^{u'} \text{ cats ate some}^{\nu'} \text{ mice}]\!]$

$$\Leftrightarrow \lambda g.\left\{ g^{\substack{\nu\mapsto y \\ u\mapsto x \\ \nu'\mapsto y' \\ u'\mapsto x'}} \middle|\; \textsc{rat}(y), \textsc{dog}(x), \textsc{ate}(x,y), \textsc{mouse}(y'), \textsc{cat}(x'), \textsc{ate}(x',y') \right\}$$

b. **Simultaneously applying mereology-based maximality operators**:

$\mathbf{M}_{u,\nu,u',\nu'} \overset{\text{def}}{=} \lambda m_{\langle s,st\rangle}.\lambda g_s.\{h \in m(g) \mid \neg\exists h' \in m(g).$

$h(u) \sqsubset h'(u) \vee h(\nu) \sqsubset h'(\nu) \vee h(u') \sqsubset h'(u') \vee h(\nu') \sqsubset h'(\nu')\}$

$\mathbf{M}_{u,\nu,u',\nu'}[\![\text{some}^u \text{ dogs ate some}^\nu \text{ rats, some}^{u'} \text{ cats ate some}^{\nu'} \text{ mice}]\!]$

$\Leftrightarrow$

$$\mathbf{M}_{u,\nu,u',\nu'}[\lambda g.\left\{ \begin{matrix} \nu\mapsto y \\ u\mapsto x \\ \nu'\mapsto y' \\ g^{u'\mapsto x'} \end{matrix} \middle| \textsc{rat}(y), \textsc{dog}(x), \textsc{ate}(x,y), \textsc{mouse}(y'), \textsc{cat}(x'), \textsc{ate}(x',y') \right\}]$$

$$\Leftrightarrow \lambda g.\left\{ \begin{matrix} \nu\mapsto y \\ u\mapsto x \\ \nu'\mapsto y' \\ g^{u'\mapsto x'} \end{matrix} \middle| \begin{matrix} y' = \Sigma y'[\textsc{mouse}(y') \wedge \textsc{ate}(x',y')] \\ x' = \Sigma x'[\textsc{cat}(x') \wedge \textsc{ate}(x',y')] \\ y = \Sigma y[\textsc{rat}(y) \wedge \textsc{ate}(x,y)] \\ x = \Sigma x[\textsc{dog}(x) \wedge \textsc{ate}(x,y)] \end{matrix} \right\}$$

c. **Comparing cardinalities**:

$$\mathbf{more}_{u,u'} \stackrel{\text{def}}{=} \lambda m_{\langle s,st\rangle}.\lambda g_s.\begin{cases} m(g) & \text{if } |g(u)| = |g(u')| + d_1 \ (d_1 > 0) \\ \varnothing & \text{otherwise} \end{cases}$$

$$\mathbf{more}_{\nu,\nu'} \stackrel{\text{def}}{=} \lambda m_{\langle s,st\rangle}.\lambda g_s.\begin{cases} m(g) & \text{if } |g(\nu)| = |g(\nu')| + d_2 \ (d_2 > 0) \\ \varnothing & \text{otherwise} \end{cases}$$

$$\mathbf{more}_{u,u'}[\mathbf{more}_{\nu,\nu'}[\mathbf{M}_{u,\nu,u',\nu'}[\![\text{some}^{u} \text{ dogs ate some}^{\nu} \text{ rats, some}^{u'} \text{ cats ate some}^{\nu'} \text{ mice}]\!]]]$$

$$\Leftrightarrow \lambda g.\left\{ \begin{matrix} \nu\mapsto y \\ u\mapsto x \\ \nu'\mapsto y' \\ g^{u'\mapsto x'} \end{matrix} \middle| \begin{matrix} y' = \Sigma y'[\textsc{mouse}(y') \wedge \textsc{ate}(x',y')] \\ x' = \Sigma x'[\textsc{cat}(x') \wedge \textsc{ate}(x',y')] \\ y = \Sigma y[\textsc{rat}(y) \wedge \textsc{ate}(x,y)] \\ x = \Sigma x[\textsc{dog}(x) \wedge \textsc{ate}(x,y)] \end{matrix} \right\}, \text{ if } \begin{matrix} |g(u)| = |g(u')| + d_1 \\ |g(\nu)| = |g(\nu')| + d_2 \end{matrix}$$

Multi-head comparatives vs. cumulative-reading sentences Just like the analysis in (54), the derivation in (55) starts with the introduction of drefs and their restrictions (see (55a), which is parallel to (54a)). The mereology-based maximality operators pick out the maximal plural individuals that satisfy all the relevant restrictions (see (55b), which is parallel to (54b), cf. Section 4.4). Finally, as shown in (55c), $\mathbf{more}_{u,u'}$ and $\mathbf{more}_{\nu,\nu'}$ convey two comparisons: (i) the cardinality of the maximal dog-sum is an increase based on the cardinality of the maximal cat-sum; (ii) the cardinality of the maximal rat-sum is an increase based on the cardinality of the maximal mouse-sum.

Multi-head comparatives vs. ordinary comparatives Just like in an ordinary comparative, the meaning of the *than*-clause in a multi-head comparative amounts to the most informative true answer to a corresponding degree question (here *how many cats ate how many mice*). Thus this *than*-clause provides the base values (i.e., $|g(u')|$ and $|g(\nu')|$) for the two increases, d_1 and d_2 (see also (52b) in Section 4.1). Then also just like in an ordinary comparative, in the matrix clause, the cardinalities of the relevant dog-sum and rat-sum are computed from the increases and the base values provided by the

than-clause (see also (52c) in Section 4.1).

It is worth noting that in a multi-head comparative like (2)/(55), the use of *more* in *more dogs* and *more rats* includes two parts: (i) comparative morpheme and (ii) gradable adjective *many* (see (56c)). Thus a multi-head comparative like (2)/(55) is also parallel to ordinary comparatives in containing elided gradable adjectives in their *than*-clause (here two instances of *many*), which can be naturally recovered in interpretation.

(56) a. Sue is tall $\underbrace{\text{er}}_{\text{comparative morpheme: an increase}}$ $\underbrace{\text{than Mary is } \sout{\text{tall}}}_{\text{how tall Mary is}}$.

b. Sue is $\underbrace{\text{more}}_{\text{comparative morpheme: an increase}}$ intelligent $\underbrace{\text{than Bill is } \sout{\text{intelligent}}}_{\text{how intelligent Bill is}}$.

c. $\underbrace{\text{More}}_{\textit{many+-er}}$ dogs ate $\underbrace{\text{more}}_{\textit{many+-er}}$ rats $\underbrace{\text{than } \sout{m\text{-many}} \text{ cats ate } \sout{n\text{-many}} \text{ mice}}_{\text{how many cats ate how many mice}}$

4.4 Maximal informativeness based on the degree-QUD of a multi-head comparative

In Sections 4.2 and 4.3, the meaning derivation of cumulative-reading sentences or multi-head comparatives involves the simultaneous use of mereology-based maximality operators (see (54b) and (55b)). As addressed in Sections 3.2 and 3.3, mereology-based maximality constitutes a special case of degree-QUD-based maximality. Thus, after the introduction of drefs and before checking or comparing cardinalities, degree-QUD-based maximality operators can be used to pick out the relativized definite drefs (see (57)).

(57) $\mathbf{M}_{u_1,u_2,\ldots} \stackrel{\text{def}}{=}$

$\lambda m.\lambda g.\,\{h \in m(g) | \neg\exists h' \in m(g)\,.\,G_{\text{QUD}}(\langle h'(u_1), h'(u_2), \ldots\rangle) >_{\text{info}} G_{\text{QUD}}(\langle h(u_1), h(u_2), \ldots\rangle)\}$

a. E.g., G_{QUD} measures informativeness along the dimensions of agents and themes: $\mathbf{M}_{u_{\text{agent}},u_{\text{theme}}} \stackrel{\text{def}}{=} \lambda m.\lambda g.\{h \in m(g) \mid \neg\exists h' \in m(g)\,.$
$G_{\text{QUD}}(\langle h'(u_{\text{agent}}), h'(u_{\text{theme}})\rangle) >_{\text{info}} G_{\text{QUD}}(\langle h(u_{\text{agent}}), h(u_{\text{theme}})\rangle)\}$

With the application of $\mathbf{M}_{u_1,u_2,\ldots}$, the drefs (which are assigned to $u_1, u_2, \ldots$) that lead to the maximal informativeness in resolving a degree QUD are selected out.

The definition of $\mathbf{M}_{u_1,u_2,\ldots}$ in (57) includes an operator G_{QUD}, which takes a tuple of drefs along multiple dimensions and returns a value indicating informativeness. Thus G_{QUD} is actually a measure function. G_{QUD} is distinct from the usual measure function contained in gradable adjectives like *tall* (see HEIGHT in the definition of $[\![\text{tall}]\!]$ in (52a)) in

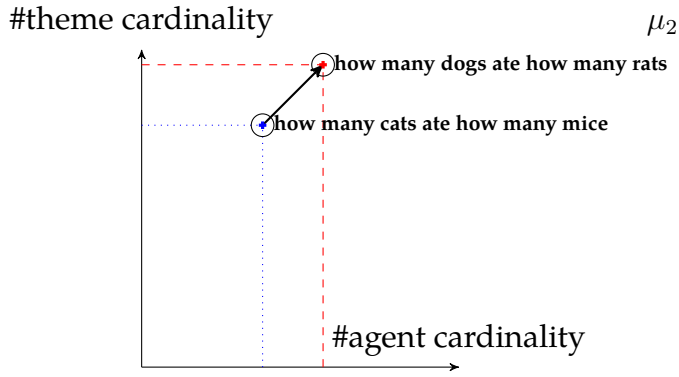

Figure 7: (2): More dogs ate more rats than cats ate mice.
Measurement of informativeness for selecting definite drefs for comparison:
$G(\langle x_{\text{agent}}, y_{\text{theme}}\rangle) = |x_{\text{agent}}| + |y_{\text{theme}}|$

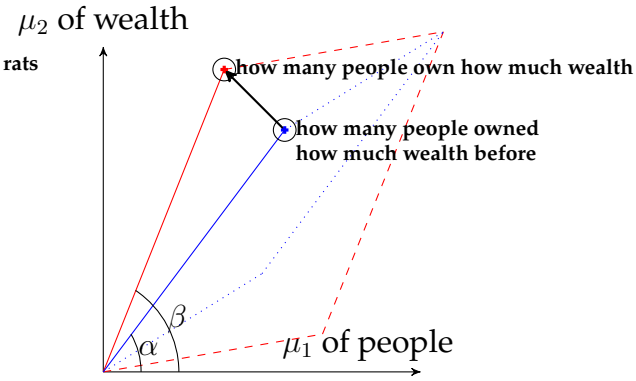

Figure 8: (59): Fewer people own more of the overall wealth.
Measurement of informativeness for selecting definite drefs for comparison:
$G(\langle x_{\text{people}}, y_{\text{wealth}}\rangle) = \frac{\mu_2(y_{\text{wealth}})}{\mu_1(x_{\text{people}})}$

that G_{QUD} uses information from potentially multiple dimensions to compress them into one overall measurement that addresses a contextually salient degree QUD.

For example, in (2)/(55) (repeated here in (74)), the underlying degree QUD is *how dogs are more successful predators than cats are*, i.e., *how successful dogs are* (which is addressed by the matrix clause) is compared with *how successful cats are* (which is addressed by the *than*-clause). Thus for both the matrix and *than*-clause, maximal informativeness amounts to mereology-based maximality along both the dimensions of agent and theme, i.e., $G(\langle x_{\text{agent}}, y_{\text{theme}}\rangle) = |x_{\text{agent}}| + |y_{\text{theme}}|$ (see Fig. 7).[7] Based on this, the degree-QUD-based maximality operators used for (2)/(55) are shown in (58a) and (58b), which are actually equivalent to mereology-based maximality operators.

(58)

More^{u} dogs ate more$^{\nu}$ rats than ~~m-many~~$^{u'}$ cats ate ~~n-many~~$^{\nu'}$ mice .

In $\mathbf{M}_{u',\nu'}$, G measures *how successful cats are* with the information of agent and theme cardinalities

In $\mathbf{M}_{u,\nu}$, G measures *how successful dogs are* with the information of agent and theme cardinalities

(= (2)/(55))

Degree QUD: *how dogs are more successful predators than cats are*

[7] Here $|x_{\text{agent}}|$ and $|y_{\text{theme}}|$ indicate the cardinality measurement along these two dimensions. The symbol '+' does not necessarily need to be understood as arithmetic addition. I use this simply to mean that the measurement of G is such that its increase depends on the increases of $|x_{\text{agent}}|$ and $|y_{\text{theme}}|$.

a. For the *than*-clause:
$\mathbf{M}_{u',\nu'} \stackrel{\text{def}}{=} \lambda m.\lambda g.\{h \in m(g) \mid \neg\exists h' \in m(g).G(\langle h'(u'), h'(\nu')\rangle) > G(\langle h(u'), h(\nu')\rangle)\}$
(the drefs that lead to the most informative true answer to *how many cats ate how many mice* will be picked out. $\rightsquigarrow$ on the plotting in Fig. 7, the **right-uppermost** point of the blue rectangle with dotted lines)

b. For the matrix clause:
$\mathbf{M}_{u,\nu} \stackrel{\text{def}}{=} \lambda m.\lambda g.\{h \in m(g) \mid \neg\exists h' \in m(g).G(\langle h'(u), h'(\nu)\rangle) > G(\langle h(u), h(\nu)\rangle)\}$
(the drefs that lead to the most informative true answer to *how many dogs ate how many rats* will be picked out. $\rightsquigarrow$ on the plotting in Fig. 7, the **right-uppermost** point of the red rectangle with dashed lines)

Below I show how to use this degree-QUD-based maximality operator (see (57)) to account for more cases of multi-head comparatives.

4.5 Multi-head comparatives with a proportion reading

As already noted by Krifka (1999), cumulative-reading sentences sometimes involve a proportion reading (see (35)). Given that multi-head comparatives are built on the base of cumulative-reading sentences, we predict that multi-head comparatives can also involve a proportion reading, and this prediction is confirmed by the existence of naturally occurring examples like (59) and (60).

(59) The trend is indisputable: Fewer people own more of the overall wealth, and fewer companies own more market share.
(A naturally occurring example:
https://www.deseret.com/opinion/2020/9/14/21436415/guest-opinion-america-capitalism-strengths-dark-side-too-far-inequality-divisiveness-wealth-gap)

(60) Fewer people own more of the land in Brazil than anywhere else in the world.
(A naturally occurring example:
https://glenmorangie.newint.org/features/2003/01/05/cutting)

Intuitively, based on our world knowledge, the most natural reading of (59) is true under a scenario where it used to be the case that 30% of the people owned 80% of the overall wealth, but now 10% of the people own 90% of the overall wealth.

Obviously, (59) addresses *how wealth distribution is more skewed nowadays than before*, i.e., *how skewed wealth distribution is now* (which is addressed by the matrix clause) is compared with *how skewed wealth distribution used to be* (which is addressed by an entirely hidden *than* clause, see (61)).

Thus, as shown in (61) and Fig. 8, at both the levels of the matrix and *than*-clause, maximal informativeness amounts to the **maximal ratio** between the measurement of the total wealth owned by some people and the measurement of their owners, i.e., $G(\langle x_{\text{people}}, y_{\text{wealth}}\rangle) = \frac{\text{the measurement of } y_{\text{wealth}}}{\text{the measurement of } x_{\text{people}}}$ (or $\frac{\mu_2(y_{\text{wealth}})}{\mu_1(x_{\text{people}})}$).

Here, for the two parts 'the measurement of 'x_{people}' and 'the measurement of y_{wealth}' (i.e., $\mu_2(y_{\text{wealth}})$ and $\mu_1(x_{\text{people}})$), the units involved in the measurements along these two dimensions, μ_1 and μ_2, are context-dependent (and often unspecified, see the two axes in Fig. 8). They can be a percentage (which leads to a proportion reading), but are not necessarily so. The units involved in μ_1 and μ_2 are actually not important (see Krifka 1999's comments on (35), which convey a similar view), but they are parallel in interpreting the matrix and the *than*-clause. Based on this measure function G, the definite drefs that correspond to the left-uppermost corner of the red and blue parallelograms in Fig. 8 are picked out and then compared (see (62) and (63)).

(61) Fewer^{u} people own more$^{\nu}$ of the overall wealth ~~than m-many$^{u'}$ people owned n-much$^{\nu'}$ of the overall wealth before~~.
Degree QUD for (59): *how wealth distribution is more skewed nowadays than before*.

a. For the *than*-clause:
$\mathbf{M}_{u',\nu'} \stackrel{\text{def}}{=} \lambda m.\lambda g.\{h \in m(g) \mid \neg\exists h' \in m(g).G(\langle h'(u'), h'(\nu')\rangle) > G(\langle h(u'), h(\nu')\rangle)\}$
(the drefs that lead to the most informative true answer to *how skewed wealth distribution used to be* will be picked out. $\rightsquigarrow$ on the plotting in Fig. 8, the **left-uppermost** corner of the blue parallelogram with dotted lines)

b. For the matrix clause:
$\mathbf{M}_{u,\nu} \stackrel{\text{def}}{=} \lambda m.\lambda g.\{h \in m(g) \mid \neg\exists h' \in m(g).G(\langle h'(u), h'(\nu)\rangle) > G(\langle h(u), h(\nu)\rangle)\}$
(the drefs that lead to the most informative true answer to *how skewed wealth distribution is* will be picked out. $\rightsquigarrow$ on the plotting in Fig. 8, the **left-uppermost** corner of the red parallelogram with dashed lines)

(62) $[\![\text{less/fewer}]\!]_d \stackrel{\text{def}}{=} d_0$ such that $d_0 \in (-\infty, 0)$ an unspecified **negative** degree
(Requirement of additivity: there is a contextually salient item serving as the base for this decrease, or negative increase. See (51b) and Zhang and Ling 2021)

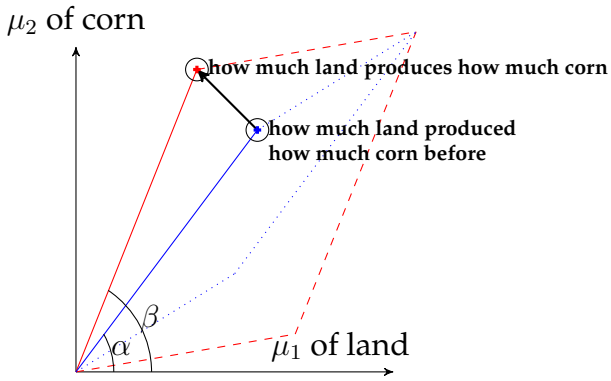

Figure 9: (3)/(64): Less land produces more corn than even before.
For the **proportion reading**, measurement of informativeness for selecting definite drefs for comparison:
$G(\langle x_{\text{land}}, y_{\text{corn}}\rangle) = \frac{\mu_2(y_{\text{corn}})}{\mu_1(x_{\text{land}})}$

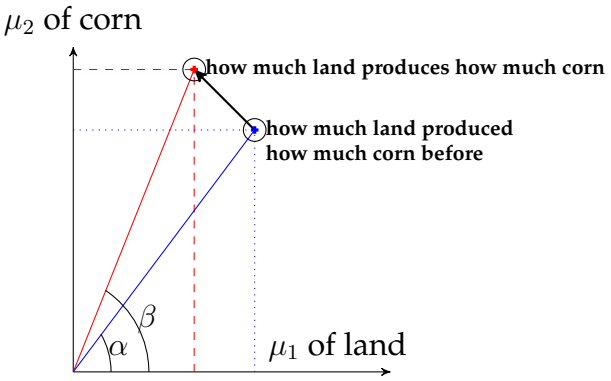

Figure 10: (3)/(64): Less land produces more corn than even before.
For the **non-proportion reading**, measurement of informativeness for selecting definite drefs for comparison:
$G(\langle x_{\text{land}}, y_{\text{corn}}\rangle) = \mu_1(x_{\text{land}}) + \mu_2(y_{\text{corn}})$

(63) Comparison tests for (59)/(61):

$$\textbf{fewer}_{u,u'} \overset{\text{def}}{=} \lambda m_{\langle s,st\rangle}.\lambda g_s.\begin{cases} m(g) & \mu_1(g(u)) = \mu_1(g(u')) + d_1 \ (d_1 < 0) \\ \varnothing & \text{otherwise} \end{cases}$$

$$\textbf{more}_{\nu,\nu'} \overset{\text{def}}{=} \lambda m_{\langle s,st\rangle}.\lambda g_s.\begin{cases} m(g) & \mu_2(g(\nu)) = \mu_2(g(\nu')) + d_2 \ (d_2 < 0) \\ \varnothing & \text{otherwise} \end{cases}$$

(In Fig. 8, the meaning of $\textbf{fewer}_{u,u'}$ is shown as the leftward direction of the arrow (i.e., a decrease along the measurement of people), and the meaning of $\textbf{more}_{\nu,\nu'}$ is shown as the upward direction of the arrow (i.e., an increase and the measurement of wealth). Overall, the left-upward direction of the arrow indicates an increase of skewness.)

(64) Less land produces more corn than ever before. (= (3))
Degree QUD: *how corn productivity is <u>better</u> than before*
The increase of productivity is indicated by a **left-upward movement** on Figs. 9 and 10: the decrease of the measurement of land and the increase of the measurement of corn.

a. With a proportion reading: $\rightsquigarrow G(\langle x_{\text{land}}, y_{\text{corn}}\rangle) = \frac{\mu_2(y_{\text{corn}})}{\mu_1(x_{\text{land}})}$
b. With a non-proportion reading: $\rightsquigarrow G(\langle x_{\text{land}}, y_{\text{corn}}\rangle) = \mu_1(x_{\text{land}}) + \mu_2(y_{\text{corn}})$

One of the core examples, (3) (repeated here as (64)), can be analyzed in the same way as (59). (3)/(64) has a proportion reading (see Fig. 9), the interpretation of the largely elided *than*-clause and the matrix clause leads to the left-uppermost corners of two parallelograms, and eventually, the sentence means a left-upward movement from the corner associated with the *than*-clause to the corner associated with the matrix clause, indicating an increase of the ratio $\frac{\mu_2(y_{\text{corn}})}{\mu_1(x_{\text{land}})}$, i.e., an increase of productivity.

(3)/(64) also has a non-proportion reading, under which the cases representing the total amount of the land where corn is/was produced and their total corn production would be selected for comparison (i.e., the two right-uppermost points of the two rectangles in Fig. 10). For this non-proportion reading, eventually, the comparison is also a left-upward movement, which indicates an increase of productivity ($\beta > \alpha$).

Obviously, the distinction between a proportion vs. a non-proportion reading of (3)/(64) affects how definite drefs are selected (see (64a) vs. (64b)). But after this step of selection, the application of comparison tests (see also (62) and (63)) is the same for both the proportion and the non-proportion reading.

4.6 Multi-head comparatives with a non-cumulative reading

Sentences like (30) (repeated here as (65)) have both a cumulative and a distributive reading. Are there multi-head comparatives based on this non-cumulative reading?

(65) Exactly three boys saw exactly five movies. (= (30))
Distributive (or non-cumulative) reading: There are exactly 3 boys such that each of them saw exactly 5 movies. (see also (30a))

Oda (2008a,b) provides Japanese data to demonstrate that in natural language, multi-head comparatives can indeed be built on a distributive reading. In (66), the use of *sorezore* ('each') in the *than*-clause and the matrix clause clearly indicates that both the matrix and the *than*-clause have a distributive reading. Without the use of *sorezore*, this sentence would be ambiguous between a cumulative and a distributive reading.

(66) San-biki-no neko-ga sorezore yon-hiki-no hatukanezumi-o tabeta **yorimo**
3-CL-GEN cat-NOM each 4-CL-GEN mouse-ACC ate THAN
(**motto**) **takusanno** inu-ga sorezore (**motto**) **takusanno** dobunezumi-o
(more) many dog-NOM each (more) many rat-ACC
tabeta.
ate

Lit. 'More dogs ate more rats each than three cats ate four mice each.'
$\rightsquigarrow$ There are 3 cats and each of them ate 4 mice. There are more than 3 dogs and each of them ate more than 3 rats. (Oda 2008b: (62) and (63))

Section 5.4 will further discuss Japanese data. In English, a sentence like (67) demonstrates the existence of multi-head comparatives with a non-cumulative reading. The most natural reading of (67) is true under a scenario where there are more boring professors than before, and each of them gave more silly lectures than before.

(67) More boring professors have given more silly lectures than ~~m-many boring professors gave n-many silly lectures~~ before. (adapted from (1))

The existence of a non-cumulative reading for (67) can be confirmed with a scenario like (68). In this scenario, the cumulative reading of (67) is certainly true. However, with regard to the QUD *whether/how education quality is worse now than before*, the utterance of (67) can be considered false and unfair. This feeling of unfairness demonstrates the existence of a non-cumulative reading for (67).

(68) Context: Suppose that in a university, it used to be the case that 100 boring professors gave in total 200 silly lectures in a semester, and now with the expansion of the university, 300 boring professors give in total 300 silly lectures.

Brasoveanu (2013) uses a silent distributivity operator δ to deal with the distributive reading of (30). As sketched out in (69), distributivity operator δ loops through all atoms of a plural individual x to check whether each atom of x saw a total of 5 movies, and then the maximal boy-sum x that satisfies this restriction is checked for its cardinality (i.e., whether it is equal to 3).

(69) Exactly three boys saw exactly five movies. **Distributive reading** of (30)
$\sigma x[\textsc{boy}(x) \wedge \delta x[\sigma y[\textsc{movie}(y) \wedge \textsc{saw}(x, y)] \wedge |y| = 5]] \wedge |x| = 3$
('Exactly 3 boys' takes scope over and gets restricted by 'exactly 5 movies')

With regard to cumulative-reading, I have shown that there is a natural connection between multi-head comparatives and corresponding sentences with multiple modified numerals. However, I argue that the non-cumulative reading of a multi-head comparative like (67) is likely not based on a corresponding derivation shown in (69).

Under the scenario shown in (70), our intuition is that (70a) is true while (70b) is

false. However, if a derivation similar to (69) is adopted, the predicted distributive reading of (70a) and (70b) should be both true, contrary to our intuition (see Fig. 11).

(70) Context: Before, there were exactly 3 boring professors in a department, and during one semester, each gave exactly 3 silly lectures. Now, this semester, there are 6 boring professors in this department, and 5 of them have given exactly 5 silly lectures, while the other one have given exactly 1 silly lecture.

a. More^{x} boring professors have given more^{y} silly lectures than ~~m-many$^{x'}$ boring professors gave n-many$^{y'}$ silly lectures~~ before.
The predicted distributive reading is true under the given scenario:
$|\sigma x[\textsc{professor}(x) \wedge \delta x[\sigma y[\textsc{lecture}(y) \wedge \textsc{gave}(x,y)] \wedge |y| > |y'|]]| >$
$|\sigma x'[\textsc{professor}(x') \wedge \delta x'[\sigma y'[\textsc{lecture}(y') \wedge \textsc{gave}(x',y')] \wedge |y| > |y'|]]|$
(i.e., the total number of professors x s.t. each atom of x has given 'more than $|y'|$' lectures exceeds the total number of professors x' s.t. each atom of x' gave $|y'|$ lectures.)

b. Fewer^{x} boring professors have given fewer^{y} silly lectures than ~~m-many$^{x'}$ boring professors gave n-many$^{y'}$ silly lectures~~ before.
The predicted distributive reading is also true under the given scenario:
$|\sigma x[\textsc{professor}(x) \wedge \delta x[\sigma y[\textsc{lecture}(y) \wedge \textsc{gave}(x,y)] \wedge |y| < |y'|]]| <$
$|\sigma x'[\textsc{professor}(x') \wedge \delta x'[\sigma y'[\textsc{lecture}(y') \wedge \textsc{gave}(x',y')] \wedge |y| < |y'|]]|$
(i.e., the total number of professors x s.t. each atom of x has given 'fewer than $|y'|$' lectures is lower than the total number of professors x' s.t. each atom of x' gave $|y'|$ lectures.)

The contrast between our intuitive judgments for (70a) and (70b) suggests that for a multi-head comparative like (67), the existence of a distributive reading à la (69), in which one comparative expression (here *more boring professors*) takes scope over and gets restricted by the other (here *more silly lectures*), is actually dubious. Thus Brasoveanu (2013)'s approach to distributivity (as shown in (69)) should not work for analyzing the non-cumulative reading of multi-head comparatives.

A relevant remark here is that the distributive analysis in (70a) and (70b) is similar to Hendriks (1994)'s analysis for multi-head comparatives shown in (21) and (22) (see Section 2.2): comparison along one dimension is embedded within (or scoped under) comparison along another dimension. Hendriks (1994) claims that the derivation for such a reading is impossible. My view is rather that although the derivation for such a

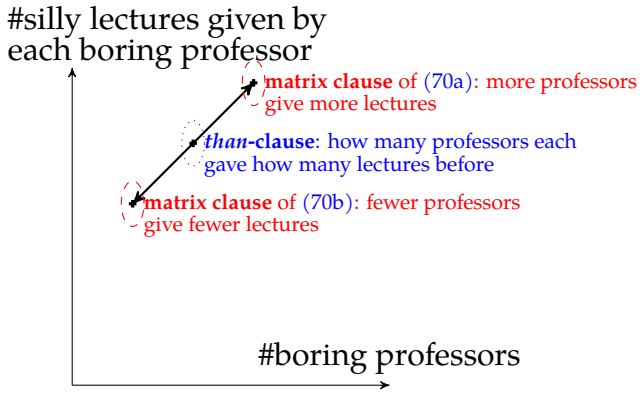

Figure 11: (67): More boring professors have given more silly lectures than before. **The analysis of (70a)/(70b)**: the two red dashed areas can co-exist, and when compared with the blue dotted area, they indicate changes of opposite directions.

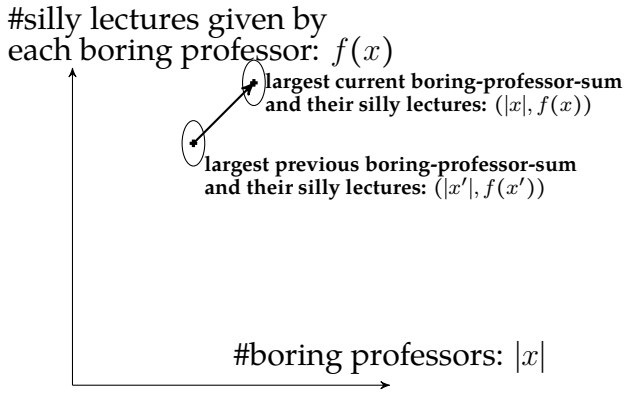

Figure 12: (67): More boring professors have given more silly lectures than before. **The current of analysis**: there are more boring professors than before (i.e., $|x| > |x'|$), and they (each) have given more silly lectures than before (i.e., $f(x) - f(x') \subseteq (0, +\infty)$).

reading is possible, this kind of reading is empirically unattested.[8] The reason is that, as pointed out by Meier (2001), 'it is hard to imagine a situation in which the comparison construction might be relevant.' The readings shown in (70a) and (70b) can be true in the same scenario, but with regard to a QUD like *how education quality is like compared to before*, they provide opposite information, suggesting that neither can resolve the QUD (see Fig. 11).

For the non-cumulative reading of (67), obviously, there is no mutual restriction between the agent and the theme (at the matrix or *than*-clause level). Furthermore, as

[8]With an interval-based implementation (see Zhang and Ling 2021), the derivation of the reading in (70a) can be sketched out like this:

(i) a. Drefs are introduced and restrictions are added onto them:
PROFESSOR(x), PROFESSOR(x'), LECTURE(y), LECTURE(y'), GAVE(x, y), GAVE(x', y')

b. Mereology-based maximality operator is applied at the *than*-clause level, selecting out the maximal professor-sum that gave lectures: $\sigma x'[\text{PROFESSOR}(x') \wedge \delta x'[\sigma y'[\text{LECTURE}(y') \wedge \text{GAVE}(x', y')]]]$. For such a maximal x', we check how many lectures each atom of x' gave (e.g., between 2 and 5 lectures) and get $|y'|$ (e.g., $|y'| = [2, 5]$).

c. Mereology-based maximality operator is applied at the matrix-clause level, selecting out the maximal professor-sum x s.t. each atom of x gave more than $|y'|$ lectures (e.g., > 5 lectures): $\sigma x[\text{PROFESSOR}(x) \wedge \delta x[\sigma y[\text{LECTURE}(y) \wedge \text{GAVE}(x, y)] \wedge |y| > |y'|]]$

d. For such a maximal x, we conduct a comparison test to check whether $|x| > |x'|$.

illustrated in (70a)/(70b) and Fig. 11, the derivation in which comparison along the cardinality of themes is used as a restriction in agent-dref selection yields an unattested reading. Thus, the only possibility to derive the non-cumulative reading for (67) is that agent-dref selection is independent from theme information.

As illustrated in Fig. 12, (67) addresses an underlying degree QUD *how education quality is worse now than before*. To address this QUD, for a boring-professor-sum x, measurement is conducted along two dimensions: (i) cardinality (i.e., $|x|$) and (ii) cadinality of silly lectures given by each atomic atom of x (i.e., $f(x)$). (67) resolves the underlying degree QUD by expressing an increase along both dimensions.

The derivation of the non-cumulative reading of (67) is shown in (71).

(71) More^{u} boring professors have given more silly lectures
than ~~m-many$^{u'}$ boring professors gave n-many silly lectures~~ before. (= (67))
the measurements of maximal $g(u')$ along two dimensions: $|g(u')|$,$|f(g(u'))|$
(each)-give-more-silly-lects$_{u,u'}$[**more**$_{u,u'}$[**M**$_{u,u'}$ [[some^{u} boring profs, some$^{u'}$ boring profs]]]
comparing along two dimensions; maximality; dref introduction

a. **Introducing drefs**: [[some^{u} boring professors, some$^{u'}$ boring professors]]
$\Leftrightarrow \lambda g.\left\{ g^{\substack{u\mapsto x\\ u'\mapsto x'}} \middle|\ \textsc{prof-c}(x), \textsc{prof-p}(x') \right\}$ (current and previous professors)

b. **Selecting out mereologically maximal drefs**:
$\mathbf{M}_{u} \stackrel{\text{def}}{=} \lambda m.\lambda g.\{h \in m(g) \mid \neg\exists h' \in m(g).G_{\text{GUD}}(h'(u)) \geq_{\text{INFO}} G_{\text{QUD}}(h(u))\}$
$\mathbf{M}_{u'} \stackrel{\text{def}}{=} \lambda m.\lambda g.\{h \in m(g) \mid \neg\exists h' \in m(g).G_{\text{GUD}}(h'(u')) \geq_{\text{INFO}} G_{\text{QUD}}(h(u'))\}$
Along the x-axis in Fig. 12, $G(x) = |x|$ $\rightsquigarrow$ mereological maximality
$\mathbf{M}_{u,u'}$[[some^{u} boring profs, some$^{u'}$ boring profs]]
$\Leftrightarrow \mathbf{M}_{u,u'}[\lambda g.\left\{ g^{\substack{u\mapsto x\\ u'\mapsto x'}} \middle|\ \begin{array}{l}\textsc{prof-c}(x)\\ \textsc{prof-p}(x')\end{array} \right\}]$
$\Leftrightarrow \lambda g.\left\{ g^{\substack{u\mapsto x\\ u'\mapsto x'}} \middle|\ \begin{array}{l} x = \Sigma x[\textsc{prof-c}(x)]\\ x' = \Sigma x'[\textsc{prof-p}(x')]\end{array} \right\}$

c. **Comparing along both dimensions**:
$\mathbf{more}_{u,u'} \stackrel{\text{def}}{=} \lambda m_{\langle s,st\rangle}.\lambda g_{s}.\begin{cases} m(g) & \text{if } |g(u)| - |g(u')| = d\ (d > 0)\\ \varnothing & \text{otherwise}\end{cases}$
(each)-give-more-silly-lects$_{u,u'} \stackrel{\text{def}}{=}$
$\lambda m_{\langle s,st\rangle}.\lambda g_{s}.\begin{cases} m(g) & \text{if } f(g(u)) - f(g(u')) \subseteq (0,+\infty)\\ \varnothing & \text{otherwise}\end{cases}$

Here $f(x) = \iota I_{\langle dt\rangle}[\forall x'' \sqsubseteq_{\text{atom}} x[|\sigma y[\textsc{lecture}(y) \wedge \textsc{gave}\ (\text{x}'', \text{y})]| \in I]]$. (i.e., the most informative interval I s.t. for each atom of x, the cardinality of all her / his lectures is within I; for interval subtraction, see Zhang and Ling 2021)

(each)-give-more-silly-lects$_{u,u'}$[**more**$_{u,u'}$[$\mathbf{M}_{u,u'}$[[some^{u} b-profs, some$^{u'}$ b-profs]]]

$$\Leftrightarrow \lambda g.\ \left\{ g^{u\mapsto x}_{u'\mapsto x'} \middle|\ \begin{array}{l} x = \Sigma x[\textsc{prof-c}(x)] \\ x' = \Sigma x'[\textsc{prof-p}(x')] \end{array} \right\}, \text{if}\ \begin{array}{l} |g(u)| - |g(u')| = d \\ f(g(u)) - f(g(u')) \subseteq (0, +\infty) \end{array}$$

(71a) shows the introduction of drefs and their restrictions: $\textsc{prof-c}(x)$ means 'sums composed of current boring professors'; $\textsc{prof-p}(x)$ means 'sums composed of previous boring professors'.[9] In (71b), maximal drefs are selected out, i.e., the definite sum including all the current boring professors is assigned to u, and the definite sum including all the previous boring professors is assigned to u'. Here the maximality operators $\mathbf{M}_u$ and $\mathbf{M}_{u'}$ are still informativeness-based (see (57)). Given that the informativeness measure function G is mereology-based, eventually, the mereologically maximal sums are selected out. (71c) shows the application of two tests on the maximal drefs. **more**$_{u,u'}$ compares the cardinality of the two maximal drefs, and **(each)-give-more-silly-lects**$_{u,u'}$ compares along the other dimension (see the y-axis in Fig. 12), i.e., the cardinality of all the lectures given by an atomic professor.

Of course, atomic professors do not necessarily give the same number of lectures. Thus, along the y-axis in Fig. 12, I use intervals to represent a range of scalar values. E.g., previously, the measurement $f(g(u'))$ results in a range $[2, 3]$ (i.e., previously, boring professors each gave between 2 and 3 silly lectures), while currently, the measurement $f(g(u))$ results in a range $[4, 5]$ (i.e., currently, boring professors each give between 4 and 5 silly lectures). Then here $f(g(u)) - f(g(u'))$ equals $[1, 3]$, a subset of $(0, +\infty)$.[10] Thus the update of applying this test does not result in an empty set, and the sentence is true in this context, which captures our intuitive judgment.

Does the analysis developed in (71) fully account for our intuition for (70a) and (70b)? This analysis certainly explains why (70b) is intuitively false under the given context: this analysis predicts that the current-professor-sum (which is associated with the matrix clause) that is under comparison includes all the 6 professors, thus this cardinality exceeds the cardinality of the previous-professor-sum (which is 3). Thus the

[9] Depending on context, we can include the part 'who give/gave lectures' in the restriction as well. Then along the y-axis in Fig. 12, the measurement would always be positive.

[10] More generally, interval subtraction is defined like this: $[y_1, y_2] - [x_1, x_2] = [y_1 - x_2, y_2 - x_1]$. See Moore (1979) for mathematical details, and see Zhang and Ling (2021) for the application of interval subtraction in the semantics of comparatives.

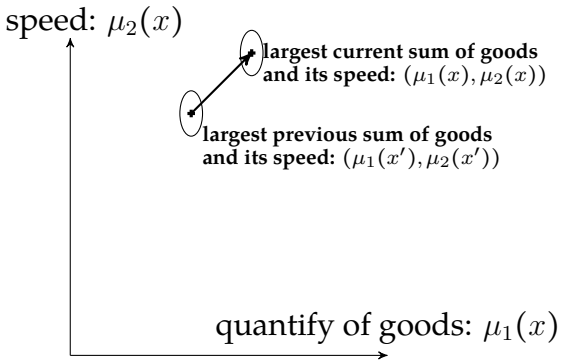

Figure 13: (28)/(72): More goods are carried faster (than before).
⇝ More goods are carried now than before, and they are carried faster.

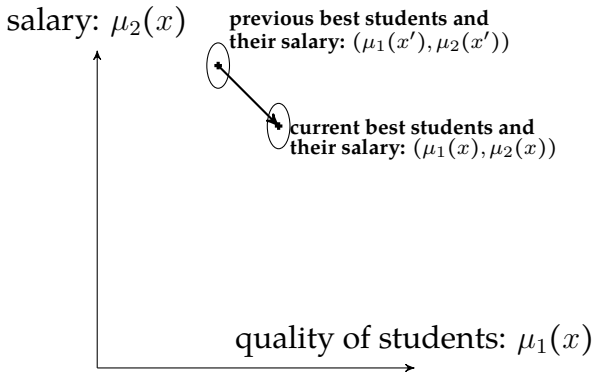

Figure 14: (73): Better students get paid worse nowadays than before.
⇝ Students are better now, but they get paid worse.

update ends with an empty set, and the sentence in (70b) is false. However, for (70a), under the current analysis, the measurement of $f(g(u))$ (which is associated with the matrix clause) is an interval $[1, 5]$, while the measurement of $f(g(u'))$ (which is associated with the *than*-clause) is $[3, 3]$, and $[1, 5] - [3, 3]$ $(= [-2, 2])$ is not a subset of the most general positive interval that represents an increase, $(0, +\infty)$. Thus (70a) is also predicted to be false, contrary to our intuition that (70a) is true under the given context.

A potential explanation for our judging (70a) true is that, for this non-cumulative reading, the measurement along the y-axis in Fig. 12 might be context-dependent: it might be 'the **average** number of lectures given by atomic professors', the **mode**, or even an impressionistic estimate. Thus, under the given context, for (70a), $f(g(u))$ is about 5, while $f(g(u'))$ is 3, and the update with the test $f(g(u)) > f(g(u'))$ does not end with an empty set, and the sentence is judged true. Our intuition for the non-cumulative reading of (67) under the context in (68) also suggests that this explanation is likely. A more detailed investigation needs experimental data and is left for future work.

Below I show two more examples of multi-head comparatives with a non-cumulative reading.

(72) More^{u} goods are carried faster ~~than m-many$^{u'}$ goods were carried n-fast~~. (= (28))

Degree QUD: *how transportation is more efficient than before*

$\underbrace{\textbf{be-carried-faster}_{u,u'}[\textbf{more}_{u,u'}}_{\text{comparing along two dimensions}}[\ \underbrace{\mathbf{M}_{u,u'}}_{\text{maximality}}\ \underbrace{[\![\text{some}^{u}\ \text{goods, some}^{u'}\ \text{goods}]\!]}_{\text{dref introduction}}]]$

$\mathbf{M}_{u,u'}$ is similar to the definition in (71b): the informativeness measurement is along the x-axis in Fig. 13, i.e., the measurement of quantity.

(72) addresses the underlying degree QUD *how transportation is more efficient than before*, and comparison is conducted along two dimensions: quantity of goods and speed (see Fig. 13). Intuitively, the selection of drefs that are goods (assigned to u and u') should be independent from speed information. To address the improvement of transportation efficiency, the total amounts of goods transported before and nowadays are taken into consideration and compared along the two dimensions (see (72)).

Eventually, (72) is true under a scenario where more goods are carried nowadays than before (i.e., $\mu_1(g(u)) > \mu_1(g(u'))$), and they are carried at a higher speed than before (i.e., $\mu_2(g(u)) > \mu_2(g(u'))$). The measurement of speed (see the y-axis of Fig. 13) might be an interval, an average, or a mode, depending on context.

(73) Context: The job market is getting worse. Previously, the best students' GPAs were around 3.6, and the month salary of their first job was between 20 k and 25 k. Now this year, the best students' GPAs are around 3.7, but the month salary of their first job is around 10 k.

Better students^{u} get paid less nowadays than ~~m-good students$^{u'}$ get paid n-much~~ before.

$\underbrace{\textbf{get-paid-less}_{u,u'}[\textbf{better}_{u,u'}}_{\text{comparing along two dimensions}}[\ \underbrace{\mathbf{M}_{u,u'}}_{\text{maximality}}\ \underbrace{[\![\text{some}^{u}\ \text{students, some}^{u'}\ \text{students}]\!]}_{\text{dref introduction}}]]$

$\mathbf{M}_{u,u'}$ is defined based on the measurement of students' quality. The best students in the past and nowadays are selected and assigned to u and u'.

(73) addresses the underlying degree QUD *how the job market is getting worse*, and comparison is conducted along the dimensions of students' quality and their salary (see Fig. 14). Again, salary information is not involved in the selection of drefs that are students (assigned to u and u'). The best students in the past and nowadays are selected and compared. (73) is true when the best students nowadays are better than the best students in the past (i.e., $\mu_1(g(u)) > \mu_1(g(u'))$), but receive lower salaries (i.e., $\mu_2(g(u)) > \mu_2(g(u'))$). In interpreting (73), vagueness might be involved in dref selection (e.g., what the threshold is for the definition of best students) and measurements along the two dimensions (e.g., intervals, averages, modes, etc.), and get resolved by context.

5 Discussion

The current work inherits insights from the existing literature, especially von Stechow (1984)'s pioneering work on (multi-head) comparatives and Brasoveanu (2013)'s formal analysis of cumulative-reading sentences. However, the current analysis takes a different perspective on the notions of comparison and informativeness, which further paves the ground for analyzing more empirical data.

5.1 Comparison: additivity vs. inequality

According to von Stechow (1984), the semantics of comparison (as expressed by comparatives) is essentially built on an inequality relation between two values, one of which is considered the comparison standard and contributed by a *than*-clause/phrase.

Thus given that a multi-head comparative expresses multiple comparisons, but at most one *than*-clause/phrase is explicitly uttered, von Stechow (1984) proposes that there is reconstruction (see (74)). Eventually, the derivation of the semantics of a multi-head comparative is based on multiple *than*-clauses.

(74) **Under von Stechow (1984)'s proposal of reconstruction** (p.43: (144)):
More dogs [than cats ate mice] ate more rats than cats ate mice. (= (2))

Under the current analysis, which follows Zhang and Ling (2021), comparison is actually the same phenomenon as additivity. Comparative morphemes like *more*/*-er* and *fewer*/*less* are similar to additive particles like (*an*)*other* and denote a positive or negative increase on the base of a contextually salient value / item. An inequality relation like '$x > y$' is considered a shorthand of '$x - y = d$' or '$x = y + d$', in which d indicates a positive increase and is expressed by *more*/*-er*. Thus comparative morphemes *more*/*-er* do not work like '>' (which is of type $\langle d, \langle dt \rangle \rangle$ and requires two arguments), and no reconstruction of another *than*-clause is needed.

Technically, the current additivity- or increase-based perspective on comparison saves the additional issue of explaining how reconstruction happens syntactically and why the reconstructed *than*-clause can never be overtly uttered.

Empirically, the semantics of a multi-head comparative is not considered two inequalities, but rather increases or decreases along two dimensions, so that it is much easier for us to consider the interplay between increases / decreases and the QUD-related interpretation behind such an interplay.

In addition, given that comparative morpheme *-er/more* denotes an increase, it can further be restricted (or modified) by degree modifiers like *slightly* and *much*. Thus a multi-head comparative like (75) resolves its underlying degree QUD with two increases that are disproportionate (*slightly* vs. *much*).

(75) Slightly better products^{u} charge much higher price than ~~m-good products$^{u'}$ charged n-high price~~ before.
Degree QUD: *how price-performance ratio is lower now than before*
Products above a certain threshold in the past and nowadays are selected for comparison, along the dimensions of quality and price (see also (73)).
$\mu_{\text{QUALITY}}(g(u)) - \mu_{\text{QUALITY}}(g(u')) \subseteq (0,$ contextual threshold of *slightly*$]$
$\mu_{\text{PRICE}}(g(u)) - \mu_{\text{PRICE}}(g(u')) \subseteq [$contextual threshold of *much*$, +\infty)$

5.2 Degree QUD-based informativeness maximality

Within Brasoveanu (2013)'s formal analysis of cumulative-reading sentences, mereologically maximal drefs that satisfy all relevant restrictions are selected out:

(76) Exactly three^{u} boys saw exactly five$^{\nu}$ movies. (= (30))
Selected drefs: all the boys who saw any movies, all the movies seen by any boys
a. Brasoveanu (2013): based on mereological maximality
b. Zhang (2023) and the current work (see also Krifka 1999): based on maximal informativeness in addressing an underlying degree QUD

The current analysis follows Krifka (1999) and Zhang (2023) and considers mereology-based maximality a special case of informativeness-based maximality: the selection of drefs is eventually based on whether they lead to maximal informativeness in addressing an underlying degree QUD. This perspective has a few consequences.

Conceptually, this degree-QUD-based perspective developed in Zhang (2023) innovates our understanding of informativeness. In the existing literature, informativeness is mainly considered based on entailment (see e.g., Fintel et al. 2014).

(77) For all x, y of type α and property ϕ of type $\langle s, \langle \alpha, t\rangle\rangle$, $x \geq_{\phi} y$ iff $\lambda w.\phi(w)(x)$ entails $\lambda w.\phi(w)(y)$. (Fintel et al. 2014: (3b))

Thus, according to Fintel et al. (2014), (78a) is more informative than (78b), because the set of worlds where the height of Mary reaches $6'2''$ is a subset of (i.e., entails) the set

of worlds where the height of Mary reaches 6 feet. However, we intuitively feel that (79a) also conveys a stronger meaning than (i.e., is also more informative than) (79b), but the entailment relation does not hold between (79a) and (79b).

(78) (78a) $\vDash$ (78b)

a. Mary is 6 feet 2 inches tall. $\lambda w.\textsc{height}(\text{Mary})(w) \geq 6'2''$

b. Mary is 6 feet tall. $\lambda w.\textsc{height}(\text{Mary})(w) \geq 6'$

(79) (79a) $\nvDash$ (79b)

a. Mary is above 6 feet tall. $\lambda w.\textsc{height}(\text{Mary})(w) > 6'$

b. Mary is between $5'9''$ and 6 feet tall. $\lambda w.\textsc{height}(\text{Mary})(w) \in [5'9'', 6']$

Under the degree-QUD-perspective on informativeness, sentences in (78) and (79) are parallel. In (78), given that $6'2''$ is a value higher than *6 feet* along the scale of height, (78a) addresses the underlying degree QUD *to what extent Mary is tall* with a higher level of informativeness than (78b). Similarly, given that *above 6 feet* indicates a value higher than *between* $5'9''$ *and 6 feet*, (79a) is more informative than (79b) in addressing the underlying degree QUD *to what extent Mary is tall*.

Another conceptual consequence is that under the current proposal, a degree-QUD-based maximality operator is actually similar to focus sensitive particles like *even*. In (80a), the uttered modified numerals invoke focus alternatives in addressing the underlying degree QUD and resolve the QUD with maximal informativeness (see also Krifka 1999; cf. Brasoveanu 2013). Similarly, in (80b), the associate of *even* (i.e., *3*) also invokes focus alternatives in addressing the underlying degree QUD and resolves the QUD with maximal informativeness. (80b) is typically not used to mean that it's unlikely for Mary or anyone else to have 3 kids. Rather, (80b) means that having 3 kids represents reaching the highest level of burden.

(80) a. At most 3% of the population own at least 70% of the land. (= (35))
Degree QUD: *how skewed wealth distribution is* (see Fig. 4)

b. Mary has even $\underline{3}_F$ kids. (see Greenberg 2018, Zhang 2022)
Degree QUD: *how heavy Mary's childcare burden is*

Empirically, this degree-QUD-based informativeness maximality naturally provides a principled way to explain how the interplay between numerals or measurements contributes to sentence interpretation. Thus the current perspective has a large empirical

coverage. For multi-head comparatives, not only those addressing a parallel change along two dimensions, but also those addressing an overall change in ratio (i.e., opposite changes along two dimensions) can be analyzed in a unified way.

Moreover, this degree-QUD-based perspective also teases apart genuine multi-head comparatives and pseudo multi-head comparatives: the former addresses their underlying degree QUD with an interplay of changes along multiple dimensions, while the latter actually addresses their underlying degree QUD only with the change along one single dimension, as illustrated in (81):[11]

(81) More schools choose a longer school year than before.
(adapted from https://www.aps.edu/news/archives/news-from-2021-2022/9-more-aps-schools-choose-a-longer-school-year)
$\rightsquigarrow$ More schools choose a longer school year than before ~~*n*-many schools chose a longer school year~~.
QUD: *how many schools choose a longer school year?*
With regard to its QUD, this comparative sentence indicates an increase along only one dimension, although apparently, there are two instances of comparative morphemes.

5.3 A related phenomenon: Comparative correlatives

Comparative correlatives are similar to multi-head comparatives in containing two comparative expressions, but comparative correlatives contain no *than*-clause. The current proposal also enables us to investigate the connection between comparative correlatives and multi-head comparatives.

Multi-head comparatives use the interplay between multi-dimensional comparisons to address an overall change (e.g., increase of consumption, increase of ratio), while comparative correlatives (see (82) and (83)) indicate how the change along one dimension (expressed by the second comparative expression) depends on the change along the other dimension (expressed by the first comparative expression).

Thus in (82), the measurement of my knowledge about my dog is an independent variable, while the measurement of my fondness for her is a dependent variable. In other words, the measurement of my fondness for her is a function on the measurement of my

[11] I thank Yusuke Yagi for reminding me of the existence of examples like (81).

knowledge about her. The changes along both dimensions share the same direction, i.e., g in (82) is a monotonically increasing function.

In (83), the value of the unit price is a function on the independent variable – the amount of purchase. The changes along these dimensions have opposite directions, i.e., g in (83) is a monotonically decreasing function.

(82) The more I know about my dog, the better I like her. (Zhang and Ling 2021)
Knowledge: x; the measurement of knowledge: $f(x)$; the measurement of fondness: $g(f(x))$ $\rightsquigarrow$ increase of $f(x)$ ~ increase of $g(f(x))$

(83) The more you buy, the lower unit price you pay.
Purchased items: x; the amount of purchase: $f(x)$; the unit price: $g(f(x))$
$\rightsquigarrow$ increase of $f(x)$ ~ decrease of $g(f(x))$

Table (84) summarizes the similarities and differences among these linguistic constructions involving multiple comparative expressions, especially with regard to how information provided by multiple comparative expressions is made use of in the overall interpretation of a sentence:

(84) Comparing linguistic constructions with multiple comparative expressions:
(Below a *change* means an increase or decrease.)

	Multi-head comparative (cumulative reading)	Multi-head comparative (non-cumulative reading)	Comparative correlative
Items involved in comparison	Co-determined by both dimensions $G(\langle x, y\rangle)$ in Figs. 7-10	Determined along one dimension $G(x)$ in Figs. 13-14	NA
Overall meaning	**Comparison** involving an interplay between the changes along $\mu_1(x)$ and $\mu_2(y)$	**Comparison** involving an interplay between the changes along $\mu_1(x)$ and $\mu_2(x)$	**Correlation**: the change along $g(f(x))$ correlates with the change along $f(x)$

5.4 Cross-linguistic data of multi-head comparatives

Cross-linguistically, Marques (2005) has discussed multi-comparatives in Portuguese, and Oda (2008a,b) has discussed multi-comparatives in Japanese. Data in these languages share a lot of similarity with English data, but there are also cross-linguistic variations. Here I address three interesting pieces of data.

First, as shown in (66) (repeated here in (85)), in Japanese multi-head comparatives, scalar values that serve as comparison standard are often overtly specified. Here in (85),

the cardinality of cats that ate mice (which is 3) and the cardinality of mice eaten by cats (which is 4) are overtly expressed in Japanese, while in English, this kind of information is derived from the interpretation of a *than*-clause (i.e., the meaning of a *than*-clause is considered equivalent to the most informative true answer to a corresponding degree question, here *how many cats ate how many mice*, see Zhang and Ling 2021 and Fleisher 2018, 2020). Oda (2008a,b) points out that this difference between English and Japanese data explains why Japanese multi-head comparatives are easier to process and interpret.

(85) San-biki-no neko-ga yon-hiki-no hatukanezumi-o tabeta **yorimo** (**motto**)
3-CL-GEN cat-NOM 4-CL-GEN mouse-ACC ate THAN (more)
takusanno inu-ga (**motto**) **takusanno** dobunezumi-o tabeta.
many dog-NOM (more) many rat-ACC ate
Lit. 'More dogs ate more rats (each) than three cats ate four mice (each).'
≈ English *more dogs are more rats than ~~m-many~~ cats ate ~~n-many~~ mice* (see (2))
⇝ Ambiguous between a cumulative and a non-cumulative reading (see (66))

Second, throughout this paper, I assume that a multi-head comparative addresses an underlying degree QUD. Marques (2005) presents a naturally occurring example in Portuguese that overtly expresses the degree QUD for comparisons along multiple dimensions. In (86), the degree QUD is about the quality of running (here *to run well*), and the comparisons that together form an interplay to address this degree QUD are along the dimensions of speed (*to run faster*) and risk level (*with less risks*). Presumably, the interpretation of (86) is similar to the non-cumulative reading of multi-head comparatives addressed in Section 4.6. Items or individuals under comparison are selected along the dimension of speed and compared along the dimension of speed and risk level. The increase along the dimension of speed and the decrease along the dimension of risk level together indicate the increase of the quality of running.

(86) Correr bem tecnicamente é correr **mais** depressa com **menos** riscos.
to-run well technically is to-run **more** fast with **less** risks
'Technically, to run well is to run faster with less risks.' (Marques 2005: p.19, (52), from a corpus available at http://www.linguateca.pt/ACDC.)

Finally, Marques (2005) uses Portuguese and English data to show that in natural language, in addition to the cumulative, non-cumulative, and ratio readings of multi-head comparatives, there is yet another subtype of comparatives that involve multiple comparisons: the comparison of differences.

In (87), the actual difference between the deficits of 2005 and 2004 exceeds the expected difference between the deficits of 2005 and 2004. In (88), the current difference between the aggressive level of Mary and her brother exceeds the previous difference.

Semantically, under the current increase or additivity-based view of comparative morphemes, Portuguese *mais* is similar to English *-er*/*more* and denotes an increase (i.e., a positive difference). It is natural that this kind of increase can be further measured and compared (see also (75)), leading to the notion of higher-order increase.

What is puzzling here is that in both Portuguese and English data, explicitly uttering $mais_{higher\text{-}order}$ *mais* in (87) or $more_{higher\text{-}order}$ *more* in (88) does not sound natural. This puzzle is left for future research.

(87) o défice de 2005 foi **mais elevado** do que o de 2004 ainda **mais** ~~[mais elevado do que o de 2004]~~ do que estava previsto
the deficit of 2005 was more high of-the what the of 2004 even more more high of-the what the of 2004 of-the what was expected
''The deficit of 2005 was higher than the 2004 one and the difference went beyond the predictions.' (Marques 2005: p. 28, (83))
$\rightsquigarrow$ the actual difference $[\mu_{\text{DEFICIT}}(2005) - \mu_{\text{DEFICIT}}(2004)]$ –
the expected difference $[\mu'_{\text{DEFICIT}}(2005) - \mu'_{\text{DEFICIT}}(2004)] > 0$

(88) Mary was always a little **more** aggressive **than** her brother. But now she's much ~~more~~ **more** aggressive **than** he is **than** she was
$\underbrace{\qquad\qquad}_{\mu_{\text{AGGRESSIVE}}(\text{Mary}) - \mu_{\text{AGGRESSIVE}}(\text{her brother})}$
$\underbrace{\text{~~more aggressive than he was~~ before.}}_{\mu'_{\text{AGGRESSIVE}}(\text{Mary}) - \mu'_{\text{AGGRESSIVE}}(\text{her brother})}$
(Marques 2005: p.21, (60), originally from Napoli 1983: (49a))
$[\mu_{\text{AGGRESSIVE}}(\text{Mary}) - \mu_{\text{AGGRESSIVE}}(\text{her brother})] -$
$[\mu'_{\text{AGGRESSIVE}}(\text{Mary}) - \mu'_{\text{AGGRESSIVE}}(\text{her brother})] \subseteq (0, \text{contextual threshold of } much)$

Both Marques (2005) and Oda (2008a,b) agree with von Stechow (1984) and Hendriks (1994), Hendriks and De Hoop (2001) that cross-linguistically, multi-head comparatives are complicated phenomena difficult to process, and often times humans lack robust intuition in interpreting them. This is not surprising, given that cumulative reading is already complicated (see Brasoveanu 2013's discussion), and multi-head comparatives are often built on a cumulative reading or involve other complex mathematical operations.

6 Conclusion

Overall, based on the inspiration from cumulative-reading sentences (and their non-cumulative reading), I have discussed and analyzed the semantics and pragmatics of multi-head comparatives. A multi-head comparative expresses comparison along two dimensions, and the interplay between changes along these two dimensions addresses an underlying degree QUD.

For a multi-head comparative with cumulative reading, the most informative cases selected for comparison are based on mutual restriction along the two dimensions involved. For a multi-head comparative with non-cumulative reading, the most informative cases selected for comparison are determined along one dimension. Eventually, for both cases, comparisons are along multiple dimensions. Thus multi-head comparatives are considered 'conceptually rather complicated (von Stechow 1984).'

The existence of this complicated phenomenon in human natural language (as well as related phenomena like comparative correlatives) sheds light on how human cognition deals with complex mathematical operations that involve measurement and comparison along multiple dimensions. In this regard, the current work echoes with existing research on intuitive mathematical computations encoded in language expressions (see also Zhang and Ling 2021 on interval subtraction, Coppock 2022 on division). The current work also suggests a new perspective on informativeness, which is not defined based on entailment, but on a degree QUD. I believe this new perspective will inspire future theoretical development and empirical investigation on informativeness.

Acknowledgments

For comments and discussions on data and ideas, I thank Duk-Ho An, Seth Cable, Heidi Harley, Jeremy Kuhn, Takeo Kurafuji, Toshiko Oda, Osamu Sawada, Alexander Williams, Yusuke Yagi, Eytan Zweig, and the audience at the symposium on 'Comparative constructions in English and other languages' at *the 41st Annual Conference of the English Linguistics Society of Japan* (University of Tokyo, 2023). I specially thank Toshiko Oda for organizing this symposium and encouraging me to work on this topic. I gratefully acknowledge financial support from NYU Shanghai. I also thank my parents and my friends (especially Jia Ling, Jinshuai Cao, Li Bai, and Quan Gan). All errors are mine.

References

Brasoveanu, Adrian. 2013. Modified numerals as post-suppositions. *Journal of Semantics* 30:155–209. https://doi.org/10.1093/jos/ffs003.

Bumford, Dylan. 2017. Split-scope definites: Relative superlatives and haddock descriptions. *Linguistics and Philosophy* 40:549–593. https://doi.org/10.1007/s10988-017-9210-2.

Charlow, Simon. 2014. On the semantics of exceptional scope. Doctoral Dissertation, NYU. https://semanticsarchive.net/Archive/2JmMWRjY/charlow-semantics-exceptional-scope-diss.pdf.

Charlow, Simon. 2017. Post-suppositions and semantic theory. https://ling.auf.net/lingbuzz/003243, *Journal of semantics* (in press).

Chomsky, Noam. 1981/1993. *Lectures on government and binding*. Foris, Dordrecht; Mouton de Gruyter. https://doi.org/10.1515/9783110884166.

Coppock, Elizabeth. 2022. Division vs. distributivity: Is *per* just like *each*? In *Semantics and Linguistic Theory*, volume 32, 384–403. https://doi.org/10.3765/salt.v1i0.5335.

Davies, Mark. 2008-. *The Corpus of Contemporary American English* (*COCA*). https://www.english-corpora.org/coca/.

Fintel, Kai Von, Danny Fox, and Sabine Iatridou. 2014. Definiteness as maximal informativeness. In *The art and craft of semantics: A festschrift for Irene Heim*, ed. Luka Crnič and Uli Sauerland, volume 2, 175 – 179. Cambridge, MA. https://semanticsarchive.net/Archive/jZiNmM4N/.

Fleisher, Nicholas. 2018. *Than* clauses as embedded questions. In *Semantics and Linguistic Theory*, volume 28, 120–140. https://doi.org/10.3765/salt.v28i0.4402.

Fleisher, Nicholas. 2020. Nominal quantifiers in *than*-clauses and degree questions. In *Syntax and Semantics Vol. 42: Interactions of Degree and Quantification*, 364–381. Brill. https://doi.org/10.1163/9789004431515_012.

Greenberg, Yael. 2010. Additivity in the domain of eventualities (or: Oliver twist's *more*). In *Proceedings of Sinn und Bedeutung*, volume 14, 151–167. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/465.

Greenberg, Yael. 2018. A revised, gradability-based semantics for even. *Natural Language Semantics* 26:51–83. https://doi.org/10.1007/s11050-017-9140-0.

Groenendijk, Jeroen, and Martin Stokhof. 1990. Two theories of dynamic semantics. In *European workshop on logics in artificial intelligence*, 55–64. Springer. https://eprints.illc.uva.nl/id/eprint/1160/1/LP-1990-14.text.pdf.

Groenendijk, Jeroen, and Martin Stokhof. 1991. Dynamic predicate logic. *Linguistics and Philosophy* 39–100. https://www.jstor.org/stable/25001418.

Haddock, Nicholas J. 1987. Incremental interpretation and combinatory categorial grammar. In *The 10th International Joint Conference on Artificial Intelligence*, volume 2, 661–663. Morgan Kaufmann Publishers Inc. https://www.ijcai.org/Proceedings/87-2/Papers/012.pdf.

Hendriks, Petra. 1994. Multiple head comparison and infinite regress. In *Proceedings of the Western Conference On Linguistics (WECOL 92)*, ed. Joel Ashmore Nevis and Vida Samiian, volume 5, 117–131. https://www.researchgate.net/publication/2814411_Multiple_Head_Comparison_and_Infinite_Regress.

Hendriks, Petra, and Helen De Hoop. 2001. Optimality theoretic semantics. *Linguistics and Philosophy* 24:1–32. http://www.jstor.org/stable/25001802.

Krifka, Manfred. 1999. At least some determiners aren't determiners. In *The semantics/pragmatics interface from different points of view*, ed. K. Turner, 257–291. Elsevier. https://semantics.uchicago.edu/kennedy/classes/w14/implicature/readings/krifka99.pdf.

Marques, Rui. 2005. Multi-headed comparatives in Portuguese. *Journal of Portuguese Linguistics* 5 (1):5–35. https://doi.org/10.5334/jpl.153.

Meier, Cécile. 2001. Multihead comparatives and result clause constructions with 'split anteced- ents'. In *Audiatur Vox Sapientiae. A Festschrift for Arnim von Stechow*, ed. Caroline Féry and Wolfgang Sternefeld, 348–371. Berlin: Akademie-Verlag. https://doi.org/10.1515/9783050080116.

Moore, Ramon E. 1979. *Methods and Applications of Interval Analysis*. SIAM. https://epubs.siam.org/doi/book/10.1137/1.9781611970906.

Napoli, Donna Jo. 1983. Comparative ellipsis: A phrase structure analysis. *Linguistic Inquiry* 14:675–694. https://www.jstor.org/stable/pdf/4178354.pdf.

Oda, Toshiko. 2008a. Degree Constructions in Japanese. Doctoral Dissertation, University of Conneticut. https://digitalcommons.lib.uconn.edu/dissertations/AAI3345201/.

Oda, Toshiko. 2008b. Multihead Comparatives in Japanese. *The Journal of Humanities and Natural Sciences* 126:3–31. http://hdl.handle.net/11150/495.

Roberts, Craige. 1996/2012. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and pragmatics* 5:6:1–69. https://doi.org/10.3765/sp.5.6.

von Stechow, Arnim. 1984. Comparing semantic theories of comparison. *Journal of Semantics* 3:1–77. https://doi.org/10.1093/jos/3.1-2.1.

Thomas, Guillaume. 2010. Incremental *more*. In *Semantics and Linguistic Theory*, volume 20, 233–250. https://doi.org/10.3765/salt.v20i0.2558.

Thomas, Guillaume. 2011. Another additive particle. In *Semantics and Linguistic Theory*, volume 21, 634–651. https://doi.org/10.3765/salt.v21i0.2589.

Zhang, Linmin. 2022. The presupposition of *even*. In *Proceedings of Semantics and Linguistic Theory*, volume 32, TBA. https://ling.auf.net/lingbuzz/006760.

Zhang, Linmin. 2023. Cumulative reading, qud, and maximal informativeness. In *Proceedings of Logic & Engineering of Natural Language Semantics* (*LENLS*), ed. D. Bekki et al., volume 19, 1–17. https://doi.org/10.1007/978-3-031-43977-3_1.

Zhang, Linmin, and Jia Ling. 2021. The semantics of comparatives: A difference-based approach. *Journal of Semantics* 38:249–303. https://doi.org/10.1093/jos/ffab003.