

# Have more people been to Paris than I have?

Interpretations of comparative illusions in Norwegian

**Lars Boye Rosseland**

Master's Thesis for the degree of  
Master of Philosophy in Linguistics



Institutt for lingvistiske, litterære og estetiske studier

Universitetet i Bergen

Norge/Norway

May 19, 2022

## Abstract

A comparative illusion is a type of sentence we seem to find acceptable, but where we seemingly are fooled by the grammar.

More people have been to Paris than I have.

People usually do not become aware of its illusory quality before some critical questions about its content is asked like: *What is actually compared?* and *What is there less of in this equation?*

Throughout previous literature on this subject it has therefore been supposed that the sentence is ungrammatical. This paper will partially challenge this assertion by suggesting elliptically adequate solutions, that will make the sentence fairly grammatical. The main problem is despite their grammaticality, they are not easily retrievable. In this paper I have called these elliptic solutions comparative deletion(CD) and comparative ellipsis(CE).

Comparative deletion: More [people]<sub>i</sub> have been to Paris than I have <sub>-i</sub>.

Comparative ellipsis: More people have [been to Paris]<sub>i</sub> than (times) I have <sub>-i</sub>.

Due to the lack of retrievability, I have also suggested two interpretations of the sentence that seems to be attractive, despite the fact that they don't fix the ellipsis. These are what Wellwood(2018) has called event comparison, called semantic coercion(SC) in the current paper. The last interpretation of the sentence *than I have* is reinterpreted into *than me*, which is called only me(OM).

In a survey, 50 participants were asked to interpret the sentence. The survey was conducted by presenting 28 variations of a comparative illusions and four different reiterations of the sentence called interpretations that respectively represent the interpretations presented above.

The goal of this study was to see if participants would significantly favour one interpretation over any other, which would suggest the participants actually interpret the sentence rather than just accepting it's grammar without any further ado. The second goal was to see if any learning effect took place during the span of the experiment. Suggesting the participants would eventually detect elliptically adequate readings of the sentence.

The results show significant results for the event comparison, but some idiosyncratic high frequencies might suggest it is more complicated. No learning effect was detected.

## Sammendrag

En komparativ illusjon er en type setning vi tilsynelatende aksepterer, men hvor grammatikken lurere oss.

Flere folk har vært i Paris enn jeg har.

Folk pleier ikke å være bevisst på setningens illusoriske kvaliteter før et par kritiske spørsmål blir stilt som *Hva er det som blir sammenlignet?* og *Hva er det mindre av i sammenligningen?*

Fra tidligere litteratur har det blitt antatt at setningen er ugrammatisk. Denne masteroppgaven vil delvis utfordre denne antagelsen med å foreslå måter ellipsen kan løses, som også gjør setningen grammatisk. Problemet med disse elliptiske løsningene er at de ikke virker til å bli oppdaget. I denne oppgaven har disse elliptiske løsningene blitt kalt komparativ sletting(CD) og komparativ sletting(CE).

Komparativ sletting: Flere [folk]<sub>i</sub> har vært i Paris enn jeg har <sub>-i</sub>.

Komparativ ellipse: Flere folk har [vært i Paris]<sub>i</sub> enn jeg har <sub>-i</sub>.

I mulig mangel av oppdelse av ellipsen, har jeg også foreslått to tolkninger av setningen som virker attraktive, tross for at de ikke tilfredsstiller kravene for at ellipsen kan løses. Disse tolkningene er hva Wellwood(2018) har kalt en begivenhetssammenligning (event comparison), forkortet som SC i denne oppgaven. Den siste tolkningen av setningen omtolker *enn jeg hartil enn meg*, som har gitt den navnet *Bare meg tolkningen*(OM).

I et eksperiment, hvor var 50 deltagere indirekte spurt hvordan de tolker setningen. Eksperimentet var utført gjennom å presentere setningen gjennom 28 variasjoner, hvor kjerneordene var erstattet sånn at setningen ble anderledes, men syntaksen lik. De fire forskjellige tolkningene av setningen ble presentert under, omskrevet slik at hver respektive tolkning blir fremhevet.

Målet med denne undersøkelsen er å se om deltagerene vil favorisere en tolkning eller ikke. Statistisk signifikante funn vil implisere at deltagerne også tolker setningen, enn å bare akseptere den uten et analytisk grunnlag. Det andre målet var å se om en læringseffekt tok sted gjennom eksperimentet. Her blir det blitt antatt at elliptiske løsninger blir oppdaget i større grad senere i eksperimentet enn tidligere.

Resultatene viser signifikante resultater i favør av begivenhetssammenligning

(SC), men resultater fra enkeltsetninger og individer tilsier at det tilsynelatende er mer kompliserende faktorer som står bak tolkningen av komparative illusjoner. Ingen læringseffekt kan påvises.

## Page of gratitude

Å skrive masteroppgave har vært en lang prosess som har vært krevende på flere måter. For det første det rent faglige. Når man holder på med ting som illusjoner og prøver å forstå hvordan de fungerer er det lett å selv bli lurt av dem om igjen og om igjen. For det andre er det alltid andre mer personlige ting som kan irritere prosessen, som gjør det å arbeide med en master utfordrende. I denne takkelisten anerkjenner jeg den hjelpen jeg har fått fra folk både faglig sett og personlig.

Jeg vil først og fremst takke alle som har vært med på eksperimentet, både i piloten og i hovedeksperimentet. Jeg vil så takke familien hjemme, som jeg har kunnet dratt hjem til for avbrekk når det har røynt som mest på.

Mine veiledere Tori Nicole Larsen og Lars-Eric Christer Johanson for å først introdusere konseptet komparativ illusjon for meg og for å ha veiledet meg gjennom ting jeg selv har funnet vanskelig å forstå, særlig når det kommer til noen av de mer kompliserte delene av statestikk.

I would like to thank Alexis Wellwood and Mario Montalbetti for their e-mail correspondances, which has helped me to understand aspects of the history of comparative illusions.

Jeg vil så til slutt takke (nei, dere er ikke glemt) Annalisa Arcidiacono og Louisa Noa Sonntag, som også deltok i pilot eksperimentet tilbake da vi tok Ling306 sammen. De har begge vært mine faste medstudenter gjennom hele masterløpet. I løpet av den tiden har vi blitt veldig nære, til et punkt hvor jeg føler jeg kan dele hva det nå enn skulle være. Det er jeg veldig takknemmelig for. Jeg er veldig glad i dere begge to. Sie wiriden immer meine Hexenhererin sein. Dein Knecht ist sehr zufrieden mit ihre Freundschaft. Ich hoffe die Herrerin ist auch zu. É la Imperatricita(ita, ita), mia amica, mia vicinal Mille grazie per la tua amicizia tra alti e bassi. Sono felice, che sono tuo Larsito.

# Contents

<b>1</b>	<b>Background</b>	<b>1</b>
1.1	Motivation for current study . . . . .	1
1.2	What is a comparative illusion? . . . . .	2
1.2.1	General outline of comparative illusions . . . . .	2
1.2.2	The illusion in Norwegian and English . . . . .	4
1.3	Thesis outline . . . . .	4
1.4	Previous research on comparative illusions . . . . .	6
1.4.1	Analysis by synthesis . . . . .	6
1.4.2	Criticisms and solutions . . . . .	9
1.5	Grammaticality and acceptability . . . . .	12
1.5.1	The use of the terms and problems therein . . . . .	12
1.5.2	Grammatical illusions . . . . .	14
1.5.3	Comparative illusion and grammaticality . . . . .	16
1.6	How to solve the ellipsis . . . . .	16
1.6.1	Comparative deletion . . . . .	17
1.6.2	Comparative ellipsis . . . . .	17
1.7	How do we interpret the illusion? . . . . .	20
1.7.1	Interpretations from Swedish and Danish students . . . . .	20
1.7.2	So why not admit that it doesn't make any sense? . . . . .	22
1.8	Pragmatic interference . . . . .	23
1.8.1	Maxims and judgements . . . . .	23
1.8.2	Pragmatic parameters intentionally excluded from the experiment . . . . .	26
1.9	Hypothesis and predictions . . . . .	27
<b>2</b>	<b>Pilot experiment</b>	<b>29</b>
2.1	The Template and Predictions . . . . .	29
2.2	The set up and execution . . . . .	32
2.3	Results . . . . .	34
2.4	local and global parsings . . . . .	37
<b>3</b>	<b>Method</b>	<b>39</b>
3.1	Experimental design . . . . .	39
3.1.1	Template . . . . .	40
3.1.2	context . . . . .	42

3.1.3	filler sentences . . . . .	42
3.1.4	The order of appearance . . . . .	43
3.1.5	Participants . . . . .	45
3.2	Experimental paradigm . . . . .	46
3.3	Ethics . . . . .	47
<b>4</b>	<b>Data and analysis</b>	<b>49</b>
4.1	Data . . . . .	49
4.1.1	Can the hypotheses be supported? . . . . .	49
4.1.2	Individual responses . . . . .	51
4.1.3	Response per sentence . . . . .	55
4.1.4	Second hypothesis: Learning effect . . . . .	62
4.2	Analysis . . . . .	66
4.2.1	What can be said about the results? . . . . .	66
4.2.2	The ungrammatical interpretation . . . . .	68
4.3	Conclusion . . . . .	69
4.3.1	Suggestions for future research . . . . .	69
<b>5</b>	<b>Appendix</b>	<b>74</b>

# 1 Background

## 1.1 Motivation for current study

Language is form of communication that emanates from natural principles of coding information, but to consistently use it, a form of systematisation is required. This is where grammar enters the frame.

We are subservient to the rules of grammar when using a certain language. This subservience shouldn't be understood in the prescriptive sense, as the term has been understood historically or through the school system or within any standardisation of a language where cognitive nuances are compromised. For instance, we used to think the split infinitive was an ungrammatical trait in English (Kroeger, 2004, 3-4), but only due to the fact that it doesn't occur in Latin, which was considered a higher form of grammar that English ought to emulate. We now understand the infinitive split to be a natural feature of English, therefore also grammatical. By these means, a descriptive understanding of grammar gives us a more naturalistic way of understanding language and has paved ways to investigate language scientifically, rather than through social judgements based on elitism or other subjective means.

This line of thinking has informed much of 20th century linguistics. In the latter part of the 20th century we've had developments in syntax and morphology (Harris, 1995, 10-73) which has given us new ways to define grammar, namely as a rule based system emanating from internal processes (Chomsky, 1965, 1980).<sup>1</sup> In a more modern definition, grammar is therefore seen as a system based on our internal knowledge. However, there seems to be other dimensions to account for when it comes to online processing of language, namely habitual thinking (Townsend and Bever, 2001). Our knowledge and habits are seen as separate, but through experiments on comparative illusions we may suggest they might not be as distinct, considering how the illusion presumably seems to be interpreted to mean something outside of its grammatical constraints.

The illusion suggests a cleft between the grammar and the parsing of language which may have consequences for the way we see syntax and human reason-

---

<sup>1</sup>I acknowledge there has been some large discussions on whether generative grammar is prescriptive or descriptive. What I am currently stating is just that the system is informed by a descriptive way of viewing language (i.e. from Bloomfield and other structuralists etc.), but not necessarily that the system itself is descriptive, which is an issue I will not further address.



ing.

So in broad strokes, what I want to achieve with the current study is first of all to give some indication as to how grammar and parsing (dis)connect to or from one another, and secondly to give a linguistic contribution to the general study of illusions, i.e. Hofstadter (1979); Bach and Poloschek (2006); Lafer-Sousa et al. (2015).

## 1.2 What is a comparative illusion?

### 1.2.1 General outline of comparative illusions

The sentence presented below (1) will seem like a normal, well functioning sentence to most people. However, by investigating it closer the meaning becomes fuzzy. That is to say; it has most likely been fuzzy all along, but the complexities are usually not initially registered. The fuzziness becomes apparent if we question the sentence in following ways; *What entities are being compared? What is there less/fewer of? and What does **have** denote?*. The concept reflected in a sentence like (1) is called a comparative illusion(CI), previously also called dead-end sentences (Christensen, 2016) or Escher sentences (Lieberman, 2004). They are called so due to their illusory quality which makes the listener/reader prematurely accept the sentence without seeming to have a coherent interpretation of its content, which may or may not come from a misreading of its form.

(1) ?More people have been to Paris than I have.

Sentence (2) is a slight modification of (1) where the comparative illusion is turned into a normal comparative sentence. This should make it apparent that it is what comes after *than* which determines its illusory quality.

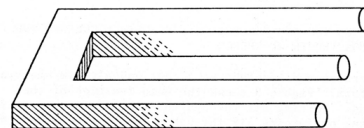
(2) More people have been to Paris than Bergen.

There ought to not be any doubt that (2) is a comparison between people in two different European cities. The amount of people whose been to Paris exceeds the amount of people whose been to Bergen, but it is not clear in (1) what the number of people in Paris exceeds.

There seem to be several aspects to the function of a comparative illusion. As I will elaborate in later sections, there are indeed ways of adequately solving the ellipsis in sentences like (1), which also makes it clear what is there is

less of. However, selective fallibility seems to play an important role in the parsing of the sentence as well, which has been the main focus in some previous research (Christensen, 2016; Wellwood et al., 2018). Selective fallibility may be the most central reason why comparative illusions are accepted overall rather than any conscious or subconscious awareness of ellipsis resolution(s).

Comparative illusions have at times been compared to other non-linguistic illusions as well. In figure 1 you see two different kinds of optical illusions. The general trick with these pictures is that they are often adequate on a local level. The smaller parts that make up the whole, i.e. the left or the right part of figure 1b). It is only when the holistic structure (the global level) is considered when they could be considered inapplicable to any reality we know. We will explore in later chapters how comparative illusions seem to build on a structure similar to the optical illusions presented.



(a) Ascending and Descending      (b) The devils tuning fork

Figure 1: McEscher(1960) "Ascending and descending" and "The devils tuning fork" have both been used as examples of optical illusions that are structurally similar to comparative illusions.

### 1.2.2 The illusion in Norwegian and English

As for comparative illusions in Norwegian, the sentence structure is approximately identical to English, including the overlapping structural and lexical ambiguity of *have*. In English *have* has two different meanings and as a consequence fill two potential roles structurally. It can either be a possessive or an auxiliary verb. Which meaning of *have* chosen, affects which ellipsis is applied, thereof structural differences as well. This ambiguity is applicable in Norwegian, as is apparent in (3).

- (3) Flere folk har vært i Paris enn jeg har  
more people have been in Paris than I have(pos/aux)  
'More people have been to Paris than I have

There's one minor difference in the English and Norwegian phrasing, namely the proposition *to/in*. It is possible such a difference makes a difference in the way people interpret the sentence, but it's doubtful since it doesn't affect the syntax or implies anything else semantically in any meaningful capacity. The application of theory where English examples are used therefore applies equally to Norwegian. It is therefore no contradiction between the English sentences used to explain the phenomenon and the Norwegian data that I will analyse.

## 1.3 Thesis outline

This paper presents two experiments on comparative illusions in Norwegian with L1 speakers as participants. In the pilot study (Rosseland et al., 2021) the acceptability of the illusion has been tested and used as a basis for the main experiment. There are no acceptability test for Norwegian prior to the pilot study. The results support acceptability of comparative illusions to a large enough extent for the question to arise whether anyone whose exposed to the sentence also makes an active interpretation together with the acceptance of the sentence, which is the main question asked in the second experiment, which I will from here on call *the main experiment*. It concerns itself with how the illusion is perceived and interpreted. This will be discussed in a twofold manner. First of all I will discuss ellipsis resolutions, meaning I will make suggestions as to how the ellipsis can be solved and as a result become grammatical sentences. Second; there are ways in which participants seem to

reanalyse the sentence by means of selective fallibility. From previous studies (Christensen, 2016; Wellwood et al., 2018) been suggested that people either coerce the meaning of the sentence to a simpler template or don't interpret the sentence at all, but merely accept it. No interpretation (meaning the participants does not find any meaning to the sentence) will not be dealt with, since interesting results can occur when it's not an option and should only be considered if none of the results are significant whatsoever.

The experiment is a survey, where participants will be exposed to comparative illusions and then given alternatives how to interpret them. Different options imply different ways the sentence is parsed. Some where the ellipsis is solved, others where the interpretation is a coercion of the form of the sentence. It has generally been assumed in previous literature that a meaning is never assigned, but is nonetheless accepted without further ado. This assertion will hopefully be challenged by the present study. A significance test will be applied to the results to see if any of the interpretations are significant by means of relative frequency. If any of the interpretations are statistically significant there will be reasons to believe the participants interpret the sentence a certain way. The participants will be exposed to a set of 28 comparative illusions with four different options, each option representative of a different interpretation of comparative illusions. After each comparative illusion is presented these interpretations, which are paraphrasing of the comparative illusion they have just witnessed will present themselves after one another. The participants will tick a box on the paraphrased sentence (from here on called interpretations) they think represent the meaning of the comparative illusion the best.

Throughout the first chapter the concept of grammatical illusion will be explained further, theories and previous experiments discussed. A set of additional theory not directly connected with comparative illusions will also be discussed and applied.

In the second chapter I will discuss the pilot experiment, which was conducted fall 2019 as a part of a master course in psycholinguistics and later released as a poster for the annual psycholinguistics conference in Flanders (Rosseland et al., 2021). The set up, results, what was figured out and how the results from the pilot makes for some of the conditions to the main experiment.

In the third chapter I will discuss the set up, design and methods applied in the

main experiment. This is where all the practical details will be discussed. The theoretical discussion in chapter 1 and the set up discussed in chapter 3 will then be synthesised in the fourth chapter, where the results and an analysis will be presented.

## 1.4 Previous research on comparative illusions

Comparative illusions was initially a subject of anecdotal interest in some academic circles. We don't have any written sources on the earliest discussions on comparative illusions, since the phenomenon was mostly colloquially discussed<sup>2</sup>. The first mention of the sentence in any literature was by Montalbetti (1984) where the phenomenon is briefly mentioned in the prologue in dedication to Herman Schultze, who introduced the concept to him. In the same paragraph it is also said that some people regarded the sentence as proof of the autonomy of syntax. Sadly, a discussion on such a topic has never been a topic properly addressed in any paper and has probably only been part of colloquial discussions among linguists at MIT in the 80's. Though the amount of studies has increased over the last decade, arguably starting with Fults and Phillips (2004) who conducted the first experiment on comparative illusion. A handful of studies has since been conducted on the topic (Wellwood et al., 2009, 2018; Phillips et al., 2011; O'Connor et al., 2013; Christensen, 2010, 2016; de Dios-Flores, 2016; Kelley, 2018; Nussbaum, 2017). In this section I will go through the papers cited above and make an overview on what has previously been said on comparative illusions and by that make it clear what we know and don't know about comparative illusions.

### 1.4.1 Analysis by synthesis

The first application of comparative illusions in any proper theoretical framework was by Townsend and Bever (2001), where the illusory workings of the sentence was explained using the *analysis by synthesis* model. This model was first introduced by Halle and Stevens (1962) for phonology. In this paper it is proposed that a transformation occurs from any phonological input we receive to the word we recognize from the phonemes. For instance, *ten percent* can be realized the following way due to assimilation rules in English:

(4) [tem pə'sent]

---

<sup>2</sup>This was confirmed by Montalbetti through an e-mail correspondence

Despite the /n/ in *ten* is realized as an [m], it is still recognized as an /n/ for the speaker because they recognize the word, which isolated will be realized as [n]. This coerces their perception of the phoneme. The synthesis produced is between the output of the phoneme and the lexical knowledge of its phonological qualities.

The same principle has later been applied to a larger set of cognitive models of perception. For Townsend and Bever, the model applies to sentence comprehension. The model suggests the speaker or listener has a set of templates they try to match with the input they receive. The knowledge of phonemes, in the case of phonology and the knowledge of different sentence structures in the case of syntax. According to Townsend and Bever, the recognition of an object through analysis by synthesis go through two stages. In the first stage we have only partially recognized the object and in the second we have properly matched the object to a mental template. In the first stage we search for appropriate templates to match and the second stage we make sure whether the template is a proper match to the input, making the comprehension of anything external a synthesis of previous knowledge (the template based on ways we have previously categorized things) and the external input.

In the case of a garden path sentence, for instance, we experience a recapitulation at the second stage as we realize that the sentence doesn't synthesise with the template picked out from stage one. We therefore have to rerun the first part of the sentence to find the proper template since a garden path sentence leads us towards the wrong template first. As in the example below, it is easier for us to initially process *raced* as an active verb before they encounter *fell*, which makes the reader re-evaluate the meaning of *raced*.

(5) The horse raced past the barn fell

If *raced* is read as passive, meaning the horse was *raced past the barn*, the word *fell* finally makes sense syntactically by being attached to the subject *horse*. This is not the initial reading, however, since we apply syntactic systematization before we have finished the sentence. There's no reason to assume *raced* is passive before we have encountered *fell*, because passive verbs are more marked than active ones and therefore not typically applicable unless we have reason to believe it is passive. According to Gorrell (1995) we tend to choose the simpler structures over the more complicated ones if encountered by an ambiguous structure, which seems to be the case concerning active and passive voice. Active voice is first of all more frequent and secondly a simpler structure

than the passive voice, giving the reader more reason for choosing the active voice, unless there's a reason to believe the verb to be passive.

The analysis by synthesis model is equally applicable when discussing comparative illusions. Similarly to how we incorrectly identify a template like active voice in sentence 5, we identify smaller sentence structures we put together to make the comparative illusion make sense. As a consequence we give the sentence the merit of sense without realizing you don't really understand the content. Sentence 6 gives suggestions to which templates are used in comprehending the comparative illusion.

- (6) a) More people have been to Paris than I
- b) People have been more to Paris than I have

The first stage of syntactic comprehension is called pseudo-syntax and is built from the associations of the listener. For longer or more complex sentences like comparative sentences it is assumed they are built from smaller sentences like the ones displayed in 6(a+b). These sentences are the templates the input is matched with. It is therefore suggested the listener will perceive of the smaller sentences it is built from 6a+b and accordingly accept the comparative illusion with two different sentences in mind. Here's how Townsend and Bever personally discussed the issue:

”Consider your intuitions as you peruse a run-on sentence like “More people have gone to Russia than I have.” At first, it seems plausible and you think you understand it. This follows from the fact that it meets a superficial template and appears to have a meaning. But then, as you reconstitute it, it does not compute, and you realize that it is not actually a sentence, and you are actually quite confused about what it really means.” (Townsend and Bever, 2001, 184)

These sentences are both coherent in their own respect and since merging of sentences is a common phenomenon in language processing, we are lead to believe that the sentence is adequate, despite the fact that we don't know what it actually means.

The comprehension of an illusion is in this sense connected with the perception of illusions in general. McEschers *ascending and descending* and *The devils tuning fork* (see figure 1) are at any stage locally adequate, but globally at

fault. What it seems like we do, however, is perceive the local parts and synthesize it with a coherent structure. The left side and the right side of *The devils tuning fork* do indeed form a coherent structure separately, but together make up an impossible structure. It takes some time for us to understand the faultiness of the global structure, because we first have to deal with the smaller templates (local structure) in order to understand the global structure of the object in question.

#### 1.4.2 Criticisms and solutions

One of the main criticisms to the analysis by synthesis approach comes from Phillips et al. (2011). It is not necessarily because the analysis by synthesis approach is wrong, but due to variables the model cannot take into account. In an experiment conducted by Wellwood et al. (2009) the participants would accept the sentence more often if the event in question is repeatable, indicating; if an event can occur several times, then the event seems to be accounted for as one of the units measured in a comparison.

- (7) a) More undergrads call their families during the week than i do (repeatable)
- b) More New Yorkers began law school this semester than I did (non-repeatable)

In the following experiment, the participants were supposed to rate the acceptability of a sentence on a Likert scale (1-7) where the repeatable predicate received a mean rate on 5.3, but the non-repeatable only 3.8. This deviation between standard errors is also statistically significant ( $F = 13.92, p < 0.001$ ).

This means the perception of the sentence is more robust than what initially believed. A repeatable event like making a phone call makes the sentence acceptable, but an event like beginning law school does not have the same acceptability rate. It is therefore believed people interpret the sentence to be about the amount of times people have been to Paris, and not the amount of people in general. Such a fallibility makes it easier to solve the ellipsis, rather than account for the whole structure. What is compared is the amount of times people have been to Paris to the amount the speaker has been, where the speaker has been more or less times to Paris than the general population. Wellwood et al. (2018) calls this interpretation of the sentence an event comparison and is considered a semantic coercion.



Apparently because we seem to prefer to count the number of times an event has occurred and possibly interpret this amount to be the subceeding amount. This is overall the overall conclusion to Wellwood et al. (2009, 2018). The illusion is made acceptable by a selective fallibility that Wellwood calls semantic coercion. To compare the amount of times people have been to Paris to the amount of times the person speaking has, is generally just a more attractive parsing than any strict adherence to the grammatical structure. What such an experiment provides, is the sensitivity of semantic content in the interpretation of the sentence, deeming a framework like analysis by synthesis too crude for explaining the intricacies of the illusion. O'Connor (2015); Leivada (2020) suggests that Townsend and Bever (2001) and Phillips et al. (2011) answer two different questions on the phenomenon. Bever on the event comparison of sentences and Philips on the shallow processing of the sentence. These approaching should therefore not be seen as opposing one another or cancel each other out, but rather pieces that both describe the larger reason for the parsing. A middle line for these approaches could be answered by the *good enough approach* by Ferreira et al. (2007). This approach will very soon be outlined.

The second variable tested for is  $\pm more$ . That is to say the use of superlative (*more*) or sublate (*less*) quantifiers in the beginning of the sentence. Wellwoods' main hypothesis for the illusions is that we depend on a shallow parsing in order to make the sentence acceptable. The use of sublate form could further support the event comparison to be the most potent interpretation, due to the fact that an event comparison becomes a lot less attractive, as will be apparent when discussing 8.

Shallow parsing is similar to the good enough approach advocated by Ferreira et al. (2007) that says there is a certain threshold for processing sentences correctly. We match the stimuli up to a certain point and then ignore the more minute features. It is not determined where exactly this threshold occurs or if everyone has the same threshold. The *good enough approach* is very general, compared to the approaches by Townsend and Bever (2001) and Phillips et al. (2011), which gives it an advantage in terms of explanation.

Wellwood has two different suggestions for how we (re)analyse the sentence: semantic coercion and syntactic reanalysis. Syntactic reanalysis is built from the idea that *more* can be interpreted as a quantifier or an adverb in the

sentence, as in *more people* (quantifier) and *been more to Paris* (adverb). Semantic coercion that we force an adequate interpretation adjacent to the inadequate semantic form presented to us: Syntactic reanalysis and semantic coercion. These ellipses will be further discussed in a later section.

This means there are no syntactically adequate readings, but that we interpret the sentence in counter to what information is presented by means of shallow parsing. These possibilities were tested by comparing the sentences to similar sentences where the quantifier was changed from superlative (*more*) into sublate form (*fewer*). The idea being that fewer cannot be interpreted as an adverb as is apparent in a sentence like:

(8) \*been fewer to Paris

Given that the sublate form was a lot less acceptable than its superlative counterpart. It can be inferred that the sentences are not syntactically reanalysed, since the results gave no significant results between superlative and sublate sentences (Wellwood et al., 2018, 559). Leivada (2020) suggests that the approaches can be integrated with the good enough approach (Ferreira et al., 2007). The good enough approach is a partial-match strategy enacted as we parse sentences. Meaning each template only requires a processing threshold to be satisfied in order for us to parse a sentence accordingly. The output in such a parsing is not necessarily the same as the input, as we don't seem to get everything right all the time. The threshold means we will be happy with the outcome, as long as it meets some minimum criteria. We otherwise don't have the time in most conversations to understand the true meaning of any sentence, which requires a slower form of reasoning. The good enough approach can both acknowledge Townsend and Bever's analysis by synthesis and Phillips et al. (2011) robustness claim since in both cases it deals with the inability to construct the sentence properly and may suggest different aspects of how we fail to recognize the illusion of the sentence.

Christensen (2016) did a similar experiment in Danish, which has a slightly different set of quantitative lexemes than English. *Flere* (more) cannot be reinterpreted as an adverb and is therefore a significant way of testing whether the theory for syntactic reanalysis is a valid parsing and to strengthen Wellwood's hypothesis. The results support acceptability for grammatical illusions, which means case for syntactic reanalysis becomes weaker. The same experiment included the same variables with quantifiers as Wellwood (*more and fewer*) used, which didn't receive any significant results, as was anticipated.

## **1.5 Grammaticality and acceptability**

The entire concept of an illusion in linguistics is the fallibility of the structure. To recognize the fallibility of the structure, it is required to have a good foundation of the system of grammar in order to clearly tell where the structure is violated. In contrast, for a sentence to be categorized as an illusion it also has to be considered acceptable. Wherein the problem of defining acceptability and grammaticality lies.

When discussing the concept of grammaticality, it is unavoidable to also talk about acceptability, and vice versa, since these concepts intermingle to the extent that the definition of one relies on the definition of the other. Sentences are for the most time grammatical and acceptable, structurally well-formed, which complements the comprehension of participants. Leivada and Westergaard (2020) has categorized comparative illusions as an acceptable, yet ungrammatical.

Acceptability and grammaticality has been defined differently within different frameworks of linguistics. In generative literature it has been the phrase structure rules that often determine grammaticality, whereas in more functionalist literature the concept of grammaticality and acceptability are considered interchangeable. For reasons pointed out above, this interchangeableness does not satisfy to the purpose I plan to use these terms. The sentences are, after all, generally acceptable. The grammaticality must be judged using different parameters. In the main experiment I have categorised different kinds of interpretations as applicable to an adequate grammatical structure and not applicable. I will therefore spend this section what I mean with the terms acceptability and grammaticality.

### **1.5.1 The use of the terms and problems therein**

The terms how they have come to be used in modern linguistics was first used by Chomsky (1965) where the terms are respectively connected to performativity and competence. The competence being the speakers knowledge of a grammar and performativity the system put into action. What differentiates the latter from the former system is extra linguistic features like "memory limitations, distractions, shifts of attention and interest, and errors" (ibid. p.11).

In other words, grammar is assumed to mirror the linguistic knowledge of the speaker and acceptability the linguistic evaluation from the speaker (sometimes confusingly called grammatical judgment). This is where we stumble over the biggest problem in determining grammaticality. Within an experimental framework it seems like we can only account for acceptability. The linguistic knowledge of the participants is always underlying and cannot reliably be accounted for in an experiment.

”Grammaticality judgments cannot be equated with grammatical knowledge. To determine properties of the underlying system requires inferential reasoning, sometimes of a highly abstract sort.” (Grimshaw and Rosen, 1990)

Grammaticality in linguistic literature has relied on introspective judgments from the linguist. Based on the sentences deemed grammatical or not grammatical we can make out a set of rules which mirror the knowledge of speaker of language X. This comes from the early definitions of grammar in generative literature where grammar is the set of rules that the speaker is accessing with their grammaticality. Chomsky (1957) defines grammar as a set of rules by systems of form formulated in a phrase structure grammar, which is a constituent analysis of different phrases and how they can be composed into different sentences with their respective derivations (ibid, 26-27). Phrase structure grammar, as part of the generative framework is an attempt to generate the knowledge of the speaker through a set of rules the speaker must follow to make a meaningful utterance. Phrase structure has been through a number of improvements and revisions and is still to some degrees present in formal frameworks of linguistic theory to this day. The biggest question put to the methodology of this approach is whether the introspection can even be considered something separate from acceptability. It should be added that the introspective method is much more meditative than a mere judgment, but whether it reflect the underlying knowledge of the speaker/listener is a completely different question all together.

Introspection as a method of exerting linguistic knowledge has been criticized, most notably by Schütze (1996), that introspection is a reflection or an analysis of accessible content. This means introspection cannot reflect linguistic knowledge, since it is only a reflection or analysis of it and not immune to memory limitation, distractedness, shifts of attention, interest or errors of the linguist. Such a criticism leads us to the non-desirable convergence of the categories. This convergence is not satisfying, since as we have previously seen, the illusion occurs in the place where the acceptability and grammaticality

does not align. They therefore have to be separate.

When grammaticality is discussed, or lack thereof, what I will consider is constraints that block any structure to be well formed.

So by considering ungrammaticality, there has to be constraints that are violated. We therefore have to accept that there are certain rules and constraints available we call grammaticality, even if we possibly do not have any satisfying methodology of outlining it. In a best case scenario we can refute certain structures by means of violations, but not affirm grammaticality, akin to the methodology of Popper (1963). There cannot be affirmations, only refutations.

In the next section I go through a set of different grammatical illusion. Throughout it will be apparent that there are certain constraints that make the sentences ungrammatical, despite the fact that we may accept them initially before we understand the error. This is essentially how I will defend my usage of grammaticality and acceptability and their distinctiveness. In this sense, grammaticality may not go as deep as our most intrinsic linguistic knowledge, but remains a set of features we base our assertions from.

### **1.5.2 Grammatical illusions**

The concept of comparative illusions has previously been classified as a subset of grammatical illusions (Phillips et al., 2011). These are sentences that seem grammatical, but only due to selective fallibility. These illusions seem to be suggestive of certain readings because of the circumstances they act within. It has been explained that "a number of grammatical constraints have a more delayed impact on language, leading to grammatical illusions." (ibid. 156). Comparative illusions share the characteristics of seeming more attractive during initial parsing, but such a classification also assumes something about the grammaticality of the sentence. This topic will be further discussed throughout this thesis. Phillips et al. (2011) has classified a couple different kinds of grammatical illusions, some of them are mentioned and explained below. They will not be exhaustively explained, but briefly looked over in order to understand the position attributed to comparative illusions. Below I have made a few selected kinds of grammatical illusions with examples and explanations. My intention of using them is to attribute some general features of grammatical illusions.

- (9) Grammatical illusions
- a) Selective fallibility in agreement comprehension:  
"The key to the cabinets are on the table."
  - b) negative polarity items(NPI): No professor will ever say that/\*A professor will ever say that/\*A professor that no student likes will ever say that

In the case of selective fallibility in agreement the listener will fail to recognize the disagreement between the head of the clause and the verb, since the approximate noun makes for a local adequate agreement the listener will more easily recognize than the disagreement between the verb and the more distant head. The head of the clause when considering agreement between subject and verb, making the sentence acceptable, despite the fact that it violates the verb agreement.

For negative polarity item there is an imbalance between the positive and the negative conditions when certain quantifiers or expressions of equal semantic weight like *any*, *ever*, *lift a finger* or *a damn thing* is used. The sentences seem only to be licensed when it is negative, but not when positive. It is because these types of quantifiers/expressions are C-commanded by the negative marker<sup>3</sup>. There are several explanations for why these sentences seem to only be licensed when negative. The most classic explanations from Ladusaw (1979) that NPIs are created from a concept in semantics called *downward entailing contexts*. It states that if a specific statement is made it implies a more general one to be true. If the general statement is false, it implies a more specific statement is equally false. If John did not eat vegetables for breakfast it is equally false that he had kale, but if he did eat vegetables for breakfast, it is not entailed that he ate kale. Vasishth et al. (2008) tested the online processing of negative polarity items by comparing NPI sentences like the ones presented in 3b. The most important one in regards to any illusory quality is the latter one: (*A professor that no student likes will ever say that*). This last type of sentence is not c-commanded by the negation, like the first sentence is. Due to the lack of c-command, it cannot be considered grammatical, but the mere presence of a negation, no matter if it provides a license for *ever* can prove a

---

<sup>3</sup>C-command is short for constituent command and refers to a structural relation in syntax where some nodes constrain the potential outcome of other nodes due to their syntactic relation, for more information the following sources are recommended: Carnie (2021); Reinhart (1976); Chomsky (1981)

higher acceptability rate. The application of a non-licensing negation increased the acceptance rate of the sentence by 15-30% from the second sentence, where there was no negation at all. This means the mere presence of a negation will fool the participants into accepting the structure.

### 1.5.3 Comparative illusion and grammaticality

As we have seen with the different kinds of grammatical illusions is that there are ways in which we can understand their ungrammaticality. There are constraints they violate, whether it is the agreement between subject and verb, the c-command licensing from a negation attached to the head. However, if we go back to comparative illusions there are no clear violation to be seen. The grammaticality of the sentence has been discussed in previous literature. Kelley (2018) makes the argument that it is the unresolvedness of the ellipsis which is why the sentence cannot be grammatically correct. In the next section, I will go through ways in which the ellipsis of the illusion in fact can be resolved.

## 1.6 How to solve the ellipsis

As I have earlier addressed, one of the main (dys)functions of the comparative illusion is the ellipsis. I displayed in the beginning of this text that (1) does not have the easy resolutions after *than* than (2). I have included sentence (1) and (2) again as sentence (9) a) and b) for the sake of convenience.

- (10) a) More people have been to Paris than I have.  
b) More people have been to Paris than Bergen.

In 9a) it is harder to fill out what is being compared, due to the ambiguity of *have* and the unfilled ellipsis. However, that doesn't mean that comparative illusions are without any resolutions by means of ellipsis. In fact, from the research on ellipsis within comparative sentences we can find two different kinds of ellipsis that I will argue make the sentence acceptable if applied. These solutions are rules that apply for different kinds of deletion in a comparative sentences, which are called comparative deletion and comparative ellipsis (Napoli, 1983). Here is a simple set illustrating the solution of both deletion rules mentioned.

- (11) a) Comparative deletion:  
More [people]<sub>i</sub> have been to Paris than I have <sub>-i</sub>.

b) Comparative ellipsis:

More people have [been to Paris]<sub>i</sub> than (times) I have <sub>-i</sub>.

These solutions should become more clear in the two subsections below where I will discuss both of them more thorough.

### 1.6.1 Comparative deletion

A comparative deletion (CD) will delete an entire compared constituent as the head of the comparative clause is recognized and rendered superfluous in the second part of the sentence (Bresnan, 1975, 48).

- (12) a) He uttered more *homilies*<sub>i</sub> than I'd ever listened to <sub>-i</sub> in one sitting  
b) Try to be as *dispassionate*<sub>i</sub> in writing your stories as you've become <sub>-i</sub> in conducting your affairs.

In all of the examples the deleted parts refer back to the head of the comparative constituents from the main clause. The first sentence compares homilies of one particular sitting to any other times homilies has occurred from previous sittings. In the second sentence the comparison is between the degrees of dispassion in writing and in conducting affairs. Neither of these sentences need to reinstate the compared constituent when the second unit of measurement is mentioned. It is already understood from *than* that something is deleted and is what occurs as the head of the main clause.

As I have shown in (10), a comparative deletion can be applied to the comparative clause *people*, which makes the comparative illusion, which resolves the sentence by means of ellipsis. In this scenario, *have* has a possessive meaning, instead of an auxiliary meaning, so that it can refer to the amount of people I have.

### 1.6.2 Comparative ellipsis

A comparative ellipsis (CE) is an optional rule that deletes element outside of the compared constituent. As we will see a comparative ellipsis can still heavily affect how the ellipsis will be read. There are several ways CE can occur and be accounted for by a different set of rules within their respective categories (Napoli, 1983, 676). I have added some formatives to the original sentences for the sake of convenience.



- (13) a) Mary wrote more books than John did (VP - deletion)  
 b) Mary wrote more *books<sub>i</sub>* than you *think<sub>j</sub>* (Null Complement Anaphora)  
 c) Mary loves *Fellini<sub>i</sub>* more than John, *Bertolucci<sub>j</sub>* (gapping)  
 d) John would *lie<sub>i</sub>* to Sue sooner than Bill would *\_i* to Jane (pseudo gapping)  
 e) I organize *\_i* more than I actually run [her *life<sub>i</sub>*] (right node raising)

In sentence a) the VP is overall deleted and replaced by *did*. In such a case the empty category is not present, but underlying as how much of literature from the time would describe it. A similar occurrence is with b), *than you think* cannot be filled in with the head of the main clause. If we were to rewrite this sentence it would have to be similar to the very cluttered way I have recited the comparative ellipsis as options in the experiment.

- (14) The amount of books you think Mary wrote is less than what Mary actually wrote

The sentence c) is according to Napoli (1983) possibly a questionable sentence. What is deleted is the verb *love*, meaning Mary loves Fellini and John loved Bertolucci, but Mary's love is stronger than John's.

A pseudo gap(d), which is dissimilar from a VP-deletion in the sense that the object is different, meaning the new object (*To Jane*) is included, yet maintains the auxiliary verb to indicate the deletion, which makes it dissimilar to gapping. In a right node raising (e), the structure of each comparative clause run parallel to one another, but they do not share verb. *Her life* occurs at the end, rather than in the beginning of the sentence. The ellipsis will therefore occur initially and will then be resolved by the object in the second comparative clause.

I have categorised the ellipsis in 10b) as a comparative ellipsis even if it has some elements that doesn't coincide with some of the criteria I put out. A comparative ellipsis will generally not make any deletion as occurs with a CD, but will resolve it with the rules accounted for above. This makes the CE proposed in 10b) stand out as a sore thumb. However, as *been* seems to be suggestive to how the sentence ought to continue and the deleted part is not part of the comparative clause should make the ellipsis qualify to be called a comparative ellipsis based on the most basic description of the term. The deleted material in the sentence occurs outside of the comparative clause. This is true that *been to Paris* is not the item which is being compared, but it adds to the meaning of the sentence. It therefore seems to solve the ellipsis similarly to how a comparative deletion would, but what is referred back to is not the

initial comparative clause, meaning it cannot be considered a comparative deletion, but fills out the most basic criteria for a comparative ellipsis, namely that something outside of the comparative clause is what is ellipped.

The ellipsis itself is open for debate whether it should be considered acceptable or grammatical to begin with. I will mostly focus on the grammaticality aspects from here on in this section, since acceptability will be accounted for when handling the data, where a high frequency of this interpretation will in fact prove the acceptability of a CE interpretation.

The first has to do with the optionality of *time*. In 10b) I put it in parenthesis to more clearly formulate the meaning intended. I will agree the sentence can become confusing without the addition of *time*, but not that it becomes ungrammatical due to this deletion. In the sentence below *times* is optional, just like it is in 10b).

(15) I saw the bird more times than (times) I could count.

I would doubt the grammaticality of this sentence would be considered ungrammatical due to this reason. It is therefore not the rule itself which seems to be the problem with the sentence.

We then proceed to discuss the fact that the sentence compares an event to a set of individuals, as I have shown more precisely in a set below.

(16)  $People \cap Paris_{loc} > [[Paris_{loc} \cap times] \in me]$

It's not as common for comparative sentences as this one. This doesn't mean it's not something that can occur in language, and to say that something we can account for by logical formulations cannot be reformulated by language just because it doesn't occur very often sounds a bit odd. Here's a comparative sentence where this phenomenon also occurs.

(17) There are more people in this world than I have met.

I would suggest this sentence is grammatical and also fully acceptable, mostly because it is a lot more straight forward than 10b). It therefore doesn't seem to be the problem what the comparative clauses in the sentence is.

I have gone through each sentence and will conclude it might not be very acceptable (that is up to the participants in the experiment to decide), but in the grammar it's functioning. It might be the compositionality of different rules that might be puzzling. None of the rules are of the type that would typically cancel each other out. It is therefore not an option in this case, and that the puzzling element would have to be the pragmatics of the sentence.

## 1.7 How do we interpret the illusion?

As we have seen in previous section, there are in fact ways in which the illusion can be interpreted where the ellipsis is accounted for and the sentence well-formed. However, this is no guaranty for well-formed interpretations as some brief discussion on the interpretations in 1.4.2 may suggest. It therefore proves itself an interesting question whether the sentences deem themselves acceptable similarly to how they deem themselves grammatical.

### 1.7.1 Interpretations from Swedish and Danish students

Some of the most concrete data we have on interpretations of the illusion comes from (Christensen, 2016, 145) where a set of Swedish and Danish students were asked how they would interpret the sentence. The results were gathered from students participating in a course about language and cognition between 2009 and 2011 from Aarhus University, Denmark and Uppsala University, Sweden. There were in all 63 participants; 38 Danish students and 25 Swedish ones. The data was collected from essays written by the students where they were asked what they thought the sentence meant. The essays were submitted anonymously. The sentence the students were presented with was the following sentence.

- (18) Flere folk har været i Paris end jeg har  
more people AUX.perf BE.pst.Perf in Paris than 1.sg pos/aux  
"more people have been to Paris than I have"

This should give some indicator as to how the sentence is generally interpreted by people. Nonetheless, it should be noticed that the informants were all students in a psycholinguistic course and therefore has a trained eye for language than the general population. They were all given time to think over the content though an essay they handed in. This is contrary to how sentences usually are perceived; immediately perceived and put together.

In table 1, there's data from a set of Swedish and Danish students and their interpretation of the sentence. Alternative a) has the highest frequency among Danish students. It is however not a likely contender for interpretation outside of Danish, as is seen with the lower frequency in Swedish. This is most likely due to the meaning of *end*, which either mean *than* or *except*. This is particular to Danish and not present in Norwegian, where *enn* can only mean

Sentences	Danish	%	Swedish	%	Total	%
a) ... [Except me]	11	28.9%	3	12.2 %	14	22.2%
b) ... [that (just) me]	8	21.1%	7	28%	15	23,8%
c) ... [more (often) than I have]	6	15,8%	1	4%	7	11.1%
d) It doesn't make sense	3	7,9 %	7	28%	10	15,9%
e) ... [than I own]	5	13,2%	4	16%	9	14.3%
f) Other	5	13,2%	3	12%	8	12,7%
Total	38	100%	25	100%	63	100%

Table 1: The interpretation of comparative illusions by Swedish and Danish students (Christensen, 2016, 135). The results were gathered from students participating in a course about language and cognition, where they each wrote an essay about the the possible meaning that were handed in anonymously. The data is based on their individual conclusions.

*than*. Interpretation a) will therefore not be further elaborated since it is idiosyncratic for Danish. This is not a phenomenon in modern Norwegian and will therefore not be considered in this thesis.

Interpretation b) is present in the main experiment. In this case, the sentence is reinterpreted to not have a verb at the end. The *more* is additive in this context, meaning it is established *I* have already been to Paris, but there is also an additional set of people. The amount of *people* and *I* are not compared to one another, which would be an unusual comparison to make, considering *I* can only be one individual and *people* indicate more than that, meaning the sentence can be correct, but not at all informative. An additional reading is therefore preferred.

Interpretation c) is an event comparison as I have briefly outlined in section 1.4.2. The emphasis of the comparison is the amount of times the event has occurred rather than the amount of people. This is a misreading of sentence (1), since the original sentence has no quantifier concerning the event, only for the amount of people. What is interesting with this alternative is that Wellwood et al. (2018) finds this option to be one of the most vital misreadings of the sentence, yet in this dataset the interpretation is surprisingly low with only 11%. The cause of this is most likely the time the students had to look over

the sentence and reconsider the meaning of it. I would suggest this reading would likely be higher if they didn't have time meditating on the sentence.

The interpretation in d) is very much straight forward one and the most popular among Swedish students. Not as popular among Danish students, supposedly because of option a).

The next interpretation; e) is dependent on the interpretation of *have* which I have earlier discussed in section 1.5.1 is a comparative deletion. This is also the case for the other Scandinavian languages discussed in this dataset. I will not discuss f) in any capacity, since these interpretations are not accounted for. There could be more than one interpretation in this category as well, and given the overall low score it should be considered a set of uncommon interpretation(s) and therefore something I will further consider for those reasons as well. From the results we do not see any homogeneity in the data, there are different interpretations that are attractive in different languages, mostly due to idiosyncrasies I have explained above. The most surprising result is sentence b), which is structurally different from the content of the original sentence. That such a sentence has such a high score in this dataset suggests it may be overwhelmingly popular in a test with different conditions. Other than that, we cannot overcome that *it doesn't make sense* category has the highest score among Swedish students, and that it probably would've had a higher turnout if it wasn't for the Danish *end*, which may suggest that the highest turnout for a language like Norwegian would be that most people in my experiment won't accept the sentence to begin with.

### **1.7.2 So why not admit that it doesn't make any sense?**

The conclusion that the sentence does not make sense seems like an option many people apparently can agree on, but that doesn't answer why they find it acceptable. The sentence is always a puzzle, people don't seem to immediately reject this sentence, meaning there is some will to accept and give it an interpretation.

If there is good will towards the sentence, then there is likely some process of decoding the sentence as well. This reconstruction of the sentence could seem like the first stage of the analysis by synthesis approach (Bever and Townsden, 2001), namely that it is merely recognized, but not thoroughly recognized, meaning what is constructed is a pseudo-syntax and not an overall analysis formed from grammatical competence. This is however not affirmative

on any accounts. This could be assumed if none of the results are significant in any way. If any of the results are significant, then there is a reason to assume the participants do decode the sentence in a specific way, that may have something to do with a general linguistic competence. The *it doesn't make sense* option is not available in my main experiment. There are a couple of reasons for that. First of all because it is in some ways an easy option to make. It is easy to just write it off when people are unsure what they think about the sentence. This is not something we do to the extent in daily life. When in a conversation it is normal to attempt to be as cooperative as possible (Grice, 1975) and therefore assign a meaning in the best possible sense. It is believed in this case that participants have to be equally cooperative when taking the experiment. In such a situation they are forced to assign a meaning to the sentence. They are also told to do it as quickly as possible, meaning what they choose is the interpretation they find the most attractive in the moment.

It is the case that some people will likely not find the meaning in the sentence. They are told to choose the sentence which seems like the most attractive to them, even if none of the options they are faced with seem totally satisfying.

## 1.8 Pragmatic interference

### 1.8.1 Maxims and judgements

As I have discussed so far, grammar is not the *end all, be all* for judging sentences. This means we cannot just look at the competence of the speaker, but also delve into pragmatics to understand the judgements of speakers from what kind of information they are fed in the moment and how they react to it. This way of viewing performativity can to some extent be more reliable than how the term has been previously used. The errors in judgements come from the way information is being processed by the recipients.

We are usually very good at understanding the context of a sentence and connect two statements together in order to make an overall meaning coherent. This is what Grice (1975) calls the cooperative principle, where several maxims have been made to opt for any potential violations. For instance, with phrases like the following (ibid, 51):

- (19) a) I am out of petrol

b) There is a garage around the corner

Sentence 19b) could be a matter of fact statement, but it is read in relation to the first sentence, namely that person b) wants to help person a). It would otherwise infringe on the maxim *be relevant*. The thing is, we are constantly looking for ways for the maxims to not be infringed, as in the conversation in 19 where the most logical interpretation of b) would be that the garage around the corner is open and that person a) can get petrol.

Similarly, we are quick to judge if a certain word is meant as a metaphor or not or whether a statement is ironic (ibid, 53) :

(20) You are the cream of my coffee

cannot be a truism as long as the person is speaking to another person.

This cooperation is very useful to guide us through conversations, but it has its shortcomings too when it comes to some judgements in probability. The cooperative principle relies on an intuitive reading of a statement, which falters in settings where extensional readings are required. Within daily conversations and for language in general, intuitional reasoning plays the most pivotal role. It is the stimuli given which seems to be the most important.

Tversky and Kahneman (1983) made a very famous example of how we do not analyse daily events into exhaustive lists of possibilities, but rather use a limited amount of heuristics such as representativeness and availability (ibid, 3).

(21) "Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations." (ibid, 11)

We then have a couple of statements about Linda where we are supposed to judge how likely Linda is any of these.

- (22) a) Linda is active in the feminist movement. (F)
- b) Linda is a bank teller. (T)
- c) Linda is a bank teller and is active in the feminist movement. (T&F)

The example is constructed to make it seem very likely that Linda is part of the feminist movement (F) and rather unlikely that she is a bank teller (T), but the most interesting results comes from introducing the 3rd category where the least and most likely scenario are combined.

A group of 88 UBC undergraduate students were asked to rank these statements and this is the following order of most to least frequent they received. (ibid, 12)

$$(23) \quad F > T\&F > T$$

The most interesting aspect of this study is that T&F was more frequent than T, which is not the case, no matter how unlikely it is that Linda is a bank teller.

The key here is the probability of the extension law. If a rule includes an extension (i.e T + F) it will be more unlikely due to the more requirements needed for the statement to be true. Such a proposition can be expressed the following way:

$$(24) \quad P(T\&F) \geq P(F)$$

No matter how unlikely T is, it is still more unlikely than if F is included in the set (unless there's a 100% probability for F).

This type of logic is part of external reasoning and not intuitive, which is why a normal listener in a conversation will not consider it when they first stumble over such statements. People seem to be fooled by the fact that *feminist* is included in the T&F, despite the fact that the probability of Linda being a feminist has to be aligned with the probability of her being a bank teller (i.e. if there's 90% chance she's a feminist and 10% she's a bank teller it is only 9% chance she is both)

I will argue that such results would not occur if we replaced the very stimulant words like *feminist* and *bank teller* with alphanumeric slots like the ones I have already referred to (T, F, T&F). In such a scenario we will not have the same pragmatic interference we currently have.

I believe similar effects could affect comparative illusions to an extent. We imagine scenarios from the presentability of given utterance, which is usually very coherent and helpful due to the cooperative principle, but in cases like Linda it seems to backfire and in a similar way it seems to backfire in regards to comparative illusions. Some more detailed account as to why it seems to backfire has been outlined in the theories discussed in section 1.3.1 and 1.3.2.



### 1.8.2 Pragmatic parameters intentionally excluded from the experiment

Interpretations of comparative illusions are not created equal for every condition. It is believed that different conditions will make different readings favourable. This claim should also later be properly tested. In the meantime I will make a couple of assumptions regarding which readings become attractive under certain conditions. These conditions will be avoided in this experiment, since the goal is to reach a general reading of the sentence with neutral conditions.

The first parameter which is likely to bend the interpretation of the sentence a certain direction is converse pairs. Converse pairs is a relation certain lexical items may have, for instance: (Kroeger, 2018, 112)

- (25) a) Michael is my advisor  
b) I am Michael's advisee

Advisor and advisee express a certain relations between each other. We can say in a similar way that:

- (26) a) I have an advisor named Michael  
b) Michael has an advisee who is me

The possessive aspect of converse pairs are especially important when it comes to the interpretation of comparative illusions. Especially since *have* is ambiguous. With converse pairs added it seems the interpretation of *have* likely will be favoured as meaning possessive and therefore making CD the most attractive interpretation of CI, like the following sentence.

- (27) More sailors went to the deck than the captain has

This is in no capacity proven by any experimental means, but the likelihood is high, which is why I have excluded it from the current experiment, since I do not want any bias in the data towards any interpretation in particular.

The clue with converse pairs and the CD interpretation has to do with the interpretation of *have*. In any relation we can call converse pairs we can refer to the one as *having* the other. The employer has employees, the clerk has customers and so forth. Therefore it is likely a CD reading would be more frequent in such an environment than a neutral sentence, since the CD reading relies on reading *have* as a possessive where the one may possess the other in some capacity.

## 1.9 Hypothesis and predictions

The four available interpretations present in the experiment are the following: comparative ellipsis, comparative deletion, semantic coercion and the only me interpretation.

I have discussed each of these interpretations over the span of several sections throughout this first chapter. The first two (CE, CD) of these has been discussed in section 1.5.1 and 1.5.2 where I have suggested two different ways to solve the ellipsis of the sentence. The other two briefly in section 1.3.2 where I discussed the previous research on comparative illusions. They are both examples previously discussed by Wellwood et al. (2018) as the event comparison and just-me interpretation.

Wellwood (first about ellipsis and then the. I predict SC will therefore be the most frequent interpretation in the experiment, despite it does not respect the grammatical criteria of well-formedness.

So if we take out the pragmatic possibilities of affecting potential attractive parsings. From earlier experiments it seems that people are quick a getting the sentence wrong (Christensen, 2016, 142). This is why it is also assumed that the event comparison (SC) will be the most frequent variable, given that the participants in this experiment are supposed to answer the survey quickly and not overthink the option. If they were supposed to meditate on the options, the results will likely be more similar to the dataset of Swedish and Danish students that I have discussed in section 1.6.1. where an option like *it doesn't make sense* or possibly one of the grammatical options could potentially be chosen more often. It has already been mentioned that such an alternative will not be available Wellwood et al. (2018) has made a good case for the selectability of the event comparison.

However, such views can change over time. It is possible that the participants will learn during the experiment and reevaluate the sentence to a more coherent meaning, given the fact that they will be more exposed to comparative illusions and therefore find another option to be more attractive later in the experiment. The experiment will be further discussed in section 3 and results in section 4. Before that I will discuss the pilot experiment and outline the results and some of the theoretical presuppositions I am relying on. The hypotheses are the following:

- 1)  $H_0$  states that there are no difference in the frequency in answers between

the different options the participants are presented with, suggesting there are no true interpretation by the sentence considered superior to another in any capacity.

2) Another hypothesis is that the participants will choose the non-elliptic options early in the survey, but then choose elliptic ones. This is believed due to a learning effect that can occur when the participants are exposed to the sentences enough times. The elliptic sentences are, as I have previously discussed, grammatical in some capacity and should therefore be the most attractive parsing after a learning effect. before the learning effect I believe *the good enough approach* Ferreira et al. (2007) to have a stronger impact on the participants.

## 2 Pilot experiment

As I have previously stated, there are several experiments implying the acceptability of comparative illusions. However, there doesn't seem to have been any previous experiments on the acceptability in Norwegian before the pilot study, executed during the fall of 2019. The results of the experiment were presented at the annual PIF conference of 2021 (Rosseland et al., 2021).

In the pilot experiment it was proven that comparative illusions are acceptable in Norwegian, just like previous experiments has implied for English among others. The main study deals with interpretations of the sentence, where the affirmed acceptability of comparative illusions strengthen the relevance of the research question.

If the pilot did not show the illusions to be acceptable, then the research question proposed in the main study would not be as fruitful as currently stated. It is because of their acceptability that we can infer that the participants may interpret the sentence. If the illusions were not acceptable we would also not believe the participants would interpret the sentence, since people usually don't contribute meaning to sentences they don't find acceptable.

The pilot study is a reaction time experiment, where the different reaction times were used to give a better indication of how inclined different participants were towards accepting the sentence.

### 2.1 The Template and Predictions

The comparative illusion was measured using reaction times, with a control sentence seemingly similar to the illusion, but apart from the original sentence it is completely grammatical and acceptable.

(28) Many people have been more to Paris than I have

The only problem with comparing this sentence with the aforementioned comparative illusion is that the latter is one word longer than the former.

(29) More people have been to Paris than I have (9 words) Many people have been more to Paris than I have. (10 words).

This difference is a problem when reaction time is the measurement used, since the times will be different, given that the participants have one more word to

Sentence condition	Quantifier	Sentence	+ more	ellipsis
Comparative illusion	more/fewer	People have been	∅	To Paris than me.
<i>questionable</i> illusion	more/fewer	People have been	+more	To Paris than me.
Grammatically inc. sentence	Many/few	People have been	∅	To Paris than me.
Grammatically cor. sentence	Many/few	People have been	+more	To Paris than me.

Table 2: This table is generating the 8 different sentences used in the experiment. The quantifiers are either sublative or superlative, which will be discussed throughout this section. The Many/more distinction comes from the illusion and the control sentence, but if more is removed from the control sentence, the sentence becomes grammatical. If more is added to the comparative illusion we get an illusion which is assumed to be weaker, or questionably even an illusion. This was assumed due to the additional quantifier position which could potentially prohibit the parser to move the quantifier of more from the subject to the event, therefore making the illusion weaker. This table was originally made for the poster of the 2021 PIF conference (Rosseland et al., 2021) and originally made by Louisa Sonntag.

parse. This makes these sentences hard to compare in this manner. Four different sentences were instead used. Two different kinds of comparative illusions one that includes the quantifier “more” before “to Paris” and one that doesn’t. In the table below these parameters are marked as ∅ or emph+more. The remaining sentence conditions are one grammatically incorrect sentence and one grammatically correct. The use of “fewer” (færre) instead of “more” (flere) was also added, making a total of 8 different conditions within each template. The number of overall templates were 9, the number of training sentences were 6 and filler sentences 30.

- (30) a) More/fewer people have been to Paris than I have (comparative illusion)  
b) Many/few people have been to Paris than I have  
c) More/fewer people have been more to Paris than I have  
d) Many/few people have been more to Paris than I have

The experiment was designed with an assumption inspired by theories of transformational grammar that the participants move the quantifier when parsing the sentence to make it seem more correct. The sentence “more people have

been to Paris than I have” will therefore become “people have been more to Paris than I have” which is a completely adequate sentence without any of the trickery a comparative illusion is inherent to. The different illusions presented were the classic variant without the quantifier *more* after *to Paris* and the one without. What was predicted was that the illusory sentence “more people have been more to Paris than I have” would be weaker than a typical comparative illusion, since the empty space to move the quantifier “more” has been blocked by another quantifier.

The non-illusory sentence without “more” is not grammatical, because it lacks a component easily detected by the participants: “Many people have been to Paris than me”. With the quantifier “many” solely, the sentence cannot become a comparative sentence. Many participants said they felt as if a word was missing in this sentence, which is true considering how the sentence was designed. The reaction times to the different sentences will be measured through positions. Position 0 is believed to be the determining position, given the fact that the additional *more* will determine the final outcome of the sentences. During the parsing of the ungrammatical sentence it is here the participants likely will notice the missing word, it is also here they will notice an additional word in the questionable illusion. For the comparative illusion it is close to where the interpretation will occur (more .. than what?), for the latter position 1 is also of great importance.

p-6	p -5	p-4	p -3	p -2	p -1	p 0	p 1	p 2	p 3
more	people	have	been	to	Paris	than	I	Have	∅
many	people	have	been	to	Paris	more	than	I	Have

Table 3: The positioning of each word in the top and bottom row where half of them are the wrong choice (marked by an asterisk) and the numbered positions of each word in relation to 0, which is where *more* occurs. It does not occur in this sentence, meaning *than* becomes the 0 point. This table will be referred back to in the next section.

As a summary of what has been discussed, the hypotheses are the following:

The  $H_0$  hypothesis was that there are no variations between the conditions presented and the additional hypotheses that:

- 1) The reaction time of comparative illusions will be longer than the non-illusion (the grammatically correct sentence)
- 2) The grammatically incorrect sentence would be have the longest reaction time of them all.
- 3) Processing effect will be detectable at position 0 (additional quantifier  $\pm more$ )

## 2.2 The set up and execution

The pilot experiment was conducted in the psycholinguistics lab at the University of Bergen, faculty of humanities. The participants were invited to join the experiment from the university campus. They were invited to the lab where they were placed in a sound isolation room in front of a computer screen and a Cedrus RB-540 Response Box. The program used for conducting the experiment was Superlab 5 (Abboud, 2015). The program was already opened and ready when the participants entered the room.

They were instructed to go through sentences presented in a g-maze, where two different words would appear on the screen simultaneously, one top option, the other bottom. The top options were matched with the front button on the response box (the yellow button) and the bottom option to the back button on the response box (red). They were instructed to always hold each hand over each of the buttons throughout the experiment to ensure the reaction times recorded to reflect their actual reactions of each sentence they were confronted with.

It was their task to choose the word most adequate for proceeding the sentence. The options were always discreetly fit or unfit for a grammatical construction. This was achieved by choosing words of particular word classes unfit for proceeding. The initial words of each sentence unfit for proceeding were verbs in infinitive form, but without the infinitive marker (å) which makes them unfit to start a sentence in Norwegian. A similar logic using inappropriate word classes was applied in all the other positions as well, making the task itself simple to follow and the participants focused on the parsing rather than the



Figure 2: Response pad

correct options.

The g-maze was non-cumulative, meaning they could never see the words they previously encountered or read the sentence in its entirety. In this manner the participants had to rely on their own memory, which is what they do in any spoken conversation. This was to not make them self correct themselves and make the illusions too apparent, as having the chance to read the sentence over again can spoil.

The top and the bottom options were accordingly correct or incorrect 50% each, this was achieved by stacking the options in a specific pattern where any correct top choice had to be complemented by a correct bottom choice and vice versa. To distribute the correct and incorrect choices accordingly ought to have made it harder for the participants to run through the experiment only clicking the top or bottom options out of habit, due to some pattern recognition of where the correct option is placed. I have illustrated of it in the table below.



p-6	p -5	P-4	p -3	p -2	p -1	p 0	p 1	p 2
more	some*	have	bread*	to	Paris	lamb*	I	sand*
to*	people	hawk*	been	me*	feels*	than	do*	have

Table 4: The positioning of each word in the top and bottom row where half of them are the wrong choice (marked by an asterisk)

The sentences themselves were also distributed in random order, making each participant experience the sentences in a different order, except for the training sentences which were all presented in the beginning of the experiment to introduce the task for the participants and to make them understand it properly before any reaction times were assigned. The measurement of the responses was done by looking at their reaction time. If the response was quick, it would be implied the sentences were thoroughly accepted without any second thought. If the reaction time was slower than average, it is implied the participants would find the sentence suspicious somehow and therefore not be acceptable to the same extent as the rest.

## 2.3 Results

There were all in all 24 participants and 20 that we could use for the analysis. Every participant with a score less than 70% were excluded from the data set, assuming they didn't understand the experiment and should therefore not be considered representative for how comparative illusions are processed.

The data was run through lmer, which is a multi-variable tool and then tested using ANOVA.

$$\text{lmer}(\log(\text{RT}) \text{ Condition*More} + ((\text{Condition*More}) - \text{Participant}) + (1 - \text{Trial}))$$

The interaction plot shows us a crossing point where we see the significance between the  $\pm$  more for position 0 and position 1. The interaction in position 0 shows a clear significance between the conditions more and more and between the illusion and the non-illusions. Since the non-illusion without more is non-grammatical it has the highest reaction time and the non-illusion with more is perfectly grammatical it has the lowest reaction time, just as expected. These results are not reflective of position 1. The non-illusion -more has an

Condition	F(1, 61)=1.98	p=0.16 n.s.
more	F(1, 35.7)=2.48	p=0.12 n.s.
Condition x More	F(1, 31.9) = 7.34	p=0.011 *

Table 5: Results for position 0. Condition is the variance between illusion and non-illusion and the +more the variance between +more and  $\emptyset$ (see table 2).

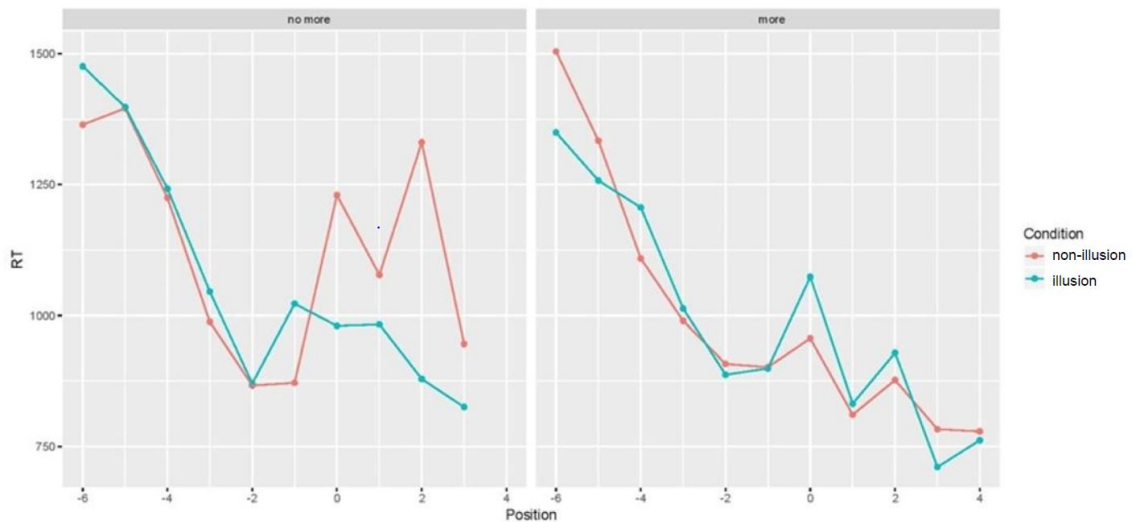
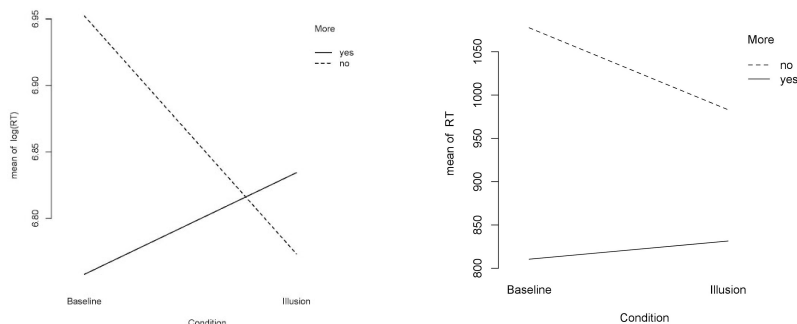


Figure 3: The reaction time for each individual position for the condition more. The blue line to the left is the comparative illusion, the red one the grammatically incorrect, the blue one to the right is the questionable illusion and the red the grammatically correct sentence

even higher reaction time, probably due to the confusion of the sentence being ungrammatical and the participants try to reconstruct the sentence. Many participants told afterwards these sentences particularly confused them and that they thought a word was missing, which is correct. It is apparent both plots are inclined to move in the same direction to one another, but the interaction plot for position 1 is not significant. As is also apparent in the p-value for each position.

This proves it right that it is at position 0 the hypothesis would be supported, since it is a decisive point for how the participants will parse the sentence. Po-

sition 1 inclines toward the same direction, but as I have previously discussed, it is position 0 which is of particular interest, therefore making the interactions of position 1 less important. As we have also seen position 1 was not significant when the condition and *more* was summarised.



(a) Interaction Position 0      (b) Interaction Position 1

Figure 4: The interaction plot for position 0 and 1

Condition	F(1,55.6)=1.02	p=0.32 n.s.
more	F(1, 46.9)=17.00	p<0.001 ***
Condition x More	F(1, 57.6)=0.78	p=0.38 n.s.

Table 6: Results for position 1. Condition is the variance between illusion and non-illusion and the +more the variance between +more and  $\emptyset$ (see table 2).

Contrary to the prediction, that the grammatically correct, grammatical illusion and the grammatically incorrect would have a gradual longer time, each

possibly significant to one another we see something different in the results. Even if the illusions have a higher reaction time in position 0 to the grammatically correct sentence, as the graph in figure 3 shows us, the variation is not significant, (p-value 0.16 n.s. and 0.12 n.s. for the illusions containing “more) meaning we cannot truly differentiate between the illusions ( $\pm more$ ) and the correct sentences. The grammatically incorrect sentences had a much higher reaction time around position 0, which verifies the fact that the more questionable a statement is, the longer the reaction time will be. These data were also significant (P=0.011 \*), meaning we can clearly distinguish the questionable qualities of the statements based on their reaction times. With that said, since the reaction time of the comparative illusions cannot be distinguished from the grammatically correct sentence, it can be concluded that comparative illusions are to be considered acceptable in Norwegian, just as it has previously been proved in other languages such as English.

## 2.4 local and global parsings

A central feature of comparative illusion, as I have previously discussed in section 1.1.1 about the topology of illusions in general is that we seem to parse the structure of a particular body locally and globally, which becomes apparent when looking at illusions. As I have shown for the devils tuning fork and the Escher staircase, the local structure is fine, but the global is faulty. If we look at the individual parts, we will not see how these illustrations cannot be representative of reality, but an overview provides us with the oddity when we look at the left and the right side of the devils tuning fork simultaneously or the entirety of the Escherian staircase and understand such a structure cannot exist in real life. A similar distinction seem to be useful in discussing sentence parsing, which became apparent when going over the data in the pilot study. In one of the options presented in the g-maze the participants were presented to grammatical and ungrammatical alternatives. The point of interest is when the participants were presented with the option *enn* (than) and *gnag* (gnaw) as you can see in the examples below. b) cannot be parsed into a coherent grammatical sentence globally, but the construction *mer + gnag* is a common expression in Norwegian that entered the data by accident. What happened was that all of the 20 participants chose *gnag* over *enn*. One reason for this could be the fact that a previous noun in the sentence was *kokker* (cooks), which is arguably semantically tied to a word like *gnag*(gnaw).

- (31) a) Få kokker har laget denne retten mer [enn] ..  
b) Få kokker har laget denne retten mer [\*gnag] ..

What is remarkable with such a finding is that the frequency in any Norwegian corpus between *mer + gnag* is remarkably smaller than *mer + enn*. What it may prove is how sentences are judged on a local level rather than a global level first and that lexical semantics may play a bigger role in parsing than the pure syntactic constraints. *Gnag* violates the global, grammatical structure, but was still chosen by all 20 participants. What is believed is that this will be reflected in the main experiment where the participants will likely chose the attractive coercive parsings over the grammatically adequate ones for similar reasons the participants chose *gnag* over *enn* in the pilot.

## 3 Method

### 3.1 Experimental design

The experiment was conducted using Machform, which is a survey maker service used by UIB. The conduction was done the following way and not in the lab, because the lab was closed due to Covid19. There were times the lab was open as well, but during that time it was too unpredictable whether the lab would be open long enough for me to finish the experiment and start the analysis. A simpler method was therefore used to conduct the experiment to my own despair. I have discussed some of these matters in the ethics section as well, where I have discussed the well being of the participants and how it is unwise to risk them potentially getting corona from being invited into the lab in person and use the same equipment as any previous participant. Despite the fact that the equipment would be cleaned between sessions.

They could only make a selection once, meaning they could not reselect an option after marking it, no multiple options were available either. Such restrictions were done in insure their replies emanated from their gut feeling and not

---

**Reven trodde han ville ta musa raskere enn hva han gjorde. \***

- Reven fikk fanget musa.
- Reven fikk ikke fanget musa.
- Reven var treigere enn musa.
- Rotta var raskere enn alle sammen.

**Beveren hadde laget en demning som var større enn hullet muldvarpen hadde gravet. \***

- Muldvarpen hadde laget et veldig stort hull.
- Beveren hadde laget en veldig stor demning.
- Demningen og hullet var like store.
- Demningen var større enn hullet.

**Flere gauper har vært på fjellet enn hjorten har. \***

- Hjorten har færre gauper enn de som var på fjellet.
- Gaupen var oftere på fjellet enn hjorten.
- Det er et større antall gauper enn ganger hjorten var på fjellet.
- Det er ikke bare hjorten som var på fjellet.

Figure 5: The comparative illusion or filler sentence presented on top and four options below each one. Each option represents a different interpretation of the sentence above. This is how it was presented to the participants, meaning there are no indication which interpretation belongs to which category. I have outlined the sentences and their categorisation in section 3.1.1.

from their long-term analytical abilities. One potential shortcoming of such an approach is misclickings, but it is overall better than not having it.

The comparative illusion is the first sentence the participants read. Different statements appear as options that the participants can choose between, all of them with similar content to the comparative illusion presented above, but varying in content. Each option reflects a different interpretation of the sentence. These interpretations are fully outlined in section 1.8. Their task was to choose the sentence they think harbour the closest meaning to the comparative illusion they've been presented. In that regard, they are actually picking the interpretation they think is the most adequate. As you can see in the figure I have presented of some of the options .. there are only the four different interpretations they can choose between. It is not possible to say that it doesn't make sense. This is defended by the cooperative principle (Grice, 1975), assuming if such sentences as those presented in the experiment ever came across in a conversation the participants would pursue a meaning to keep pace of the conversation rather than to give up. Even participants who later told me that they didn't find any of the sentences appropriate and that there seemed to be something missing from the original sentence went through and interpreted the sentence. I will argue this approach is better than to have the *it doesn't make sense* option available, because sense is something we always seek, whether it's there or not.

### 3.1.1 Template

In the section above I showed how the experiment looked like for the participants when they were presented the experiment. What it lacks is the proper categorisation of each interpretation. I have already discussed the interpretations individually, but I have not discussed how they are phrased in these sentences to be suggestive of a specific reading.

In table 7 we first see the CI (comparative illusion), which the reader should be well aware of by now, I will therefore not discuss this sentence in any further capacity. The comparative illusion presented is taken directly from the experiment and will therefore differ a bit from the usual Paris example I have otherwise used up to now.

Some of these sentences had to be somehow modified to be brief enough or

	Norwegian	English
CI	Flere gauper har vært på fjellet enn hjorten har	More lynxes has been to the mountain than the deer has
CD	Hjorten har færre gauper enn de som var på fjellet.	The deer has fewer lynxes than the ones who were on the mountain.
SC	Gaupen var oftere på fjellet enn hjorten	The lynx was more often on the mountain than the deer.
CE	Det er et større antall gauper enn ganger hjorten var på fjellet.	There's a larger amount of lynxes than times the deer was on the mountain.
OM	Det er ikke bare hjorten som var på fjellet	It was not only the deer who was on the mountain.

Table 7: The comparative illusion and the different interpretations of each sentence in Norwegian (as it was presented in the experiment) and the English translation.

to be more dissimilar from the original CI as possible. This is the case with CD. To emphasize that it is the deer who has lynxes, possessively speaking, I had to put the deer in the front of the sentence, which also meant I had to modify the quantifier from *more* to *fewer*.

SC is an event comparison. It could either mean the deer was more often on the mountain, indicating it has generally been more times.

CE had to be the largest sentence, mostly to make it grammatical and to emphasize the two items being compared, namely the amount of times the deer has been to the mountain and the amount of lynxes, which may be the least intuitive interpretation due to the strange comparison of an event and objects, which just doesn't seem as common to compare in natural language.

OM was particularly hard to rephrase. *More* needs to have an additional meaning in this case, meaning the deer has been to the mountain, but that there are additional lynxes who has also been there. I have trusted the pragmatics of the sentence presented to make the participants understand the interpretation I have intended. They were also after all told to interpret the sentence. The sentence is in all technicality analytically true in the sense that *more lynxes has been to the mountain* presupposes the deer is not the only one. However, they would have to agree that the deer actually has been to the mountain, which is not the case with a CD interpretation (the deer could own lynxes



who has been to the mountain, but could still have never been there). I will assume they will not choose this interpretation based on the analytic truth to it, but rather because they find the sentence to be the most relevant to how they personally interpret the sentence. As will be discussed in the data later on, this seems to be the case.

### 3.1.2 context

The context for the sentences in the experiments all have to do with animals found in Norwegian nature and events that fit the environment of the animals like running through the woods, jumping over a river, or to be on the top of a mountain. The two animals compared in each sentence always have some features in common, like deer and stag(Cervidae), mouse and rat(rodents) in order to not bring the attention of the participants elsewhere by focusing too much on the lexical aspect of the sentence in question. They are not always taxonomically related, but they are usually similar in size and features like  $\pm predator$  and  $\pm flight$ . They are approximately the same size or categorized in a similar category. For instance, I will not place a predator and prey in the same sentence, since it can give some pragmatic implementation like *the wolf has the sheep*. Meaning the sheep has been caught by the wolf. The implementation of different lexemes used accordingly has been discussed in section 1.7.

### 3.1.3 filler sentences

There are all in all 48 sentences: 28 comparative illusions and 20 filler sentences. The amount of filler sentences is chosen to cover 1/3 of the total amount of sentences. It is important for the filler sentences to be challenging, albeit not too challenging. If the filler sentences are too easy they will more likely detect which sentences are the filler sentences and which ones are the sentences they are being tested on. The filler sentences were therefore carefully crafted to look a bit like a comparative illusion i.e. the sentence is a comparative sentence, but not an illusion. For instance:

(32) More ducks bathed in the pond than the amount of pigeons in the park.

In such a scenario I would make options like:

(33) There were more ducks than pigeons

and then vice versa:

(34) There were more pigeons than ducks

In the filler sentences there was always one correct answer among the four, this was done in order to not exhaust the participants with ambiguities.

Another thing I consciously went through when conducting the experiment was the use of aspect. The sentence was imperfective and the options perfective. The reason for the use of aspect is two folded. First of all, it was used to not make the sentences too much alike its options below and second to make it clearer that the options are more like statements about the sentence that are either true or not true. The sentence presents a scenario and is imperfect as the reader reads it. The statements under it are in perfect as we are now only referring to the scenario presented in the sentence and should therefore be seen as an event that took part in the past.

#### **3.1.4 The order of appearance**

Another pattern participants could be prone to notice is the order in which the sentences are presented in regards to how often a comparative illusion and a filler sentence appears. They were therefore randomly scattered without any coherent logic to it. If a comparative illusion i.e. arrived in every second sentence it is likely the participants could more easily be trained to recognize a pattern. It is likely they would therefore choose the most attractive solution from an early parsing and then insert the interpretation without further effort throughout the experiment. Another way to keep the participants from recognizing the patterns was to also scatter the order each answer appeared. If each answer always appear in the same order it will be easier to recognize the phrasing of each option and therefore pick the option they have already recognized early on to be the most adequate and therefore not picking any other option for the rest of the run without any reconsideration.

The number of comparative illusions(28) was to some extent chosen due to mathematical reasons. Given there were four different alternatives to each sentence: A, B, C and D. The number 28 makes each answer occur in every position 7 times. In order to decrease any bias towards any of the alternatives, given that some participants possibly could choose i.e. the top one or

the bottom one repeatedly if they recognize a pattern. The sentences were sorted evenly in order of appearance. There are 24 different ways ABCD can be ordered. I made a list in alphabetical order like this:

- (35) ABCD  
ABDC  
ACBD  
ACDB etc.

Yet four sentences still remained. I therefore made a list where each line was displayed and added 4 more to get the proper amount required. How I solved the problem with the 4 remaining lines was to first pick a random order, in this case ABCD, and then reverse the order (DCBA). I then proceeded to invert both of them (ABCD  $\rightarrow$  BADC) and (DCBA  $\rightarrow$  CDAB).

This left me with a system where every letter would occur in the same position an equal amount of times. A appears in the first place 7 times, but so does B,C and D. The same is true for position 2,3 and 4. A,B,C,D was then assigned values.

- (36) A : Comparative deletion  
B : semantic coercion  
C : Comparative ellipsis  
D : Only me.

The options were thus evenly distributed.

In this experiment it is not as easy to determine if the participants understood the assignment or not. For instance in the pilot experiment, there was a lower limit of 70% of correct answers. If any lower, it was expected that they didn't understand it. Because of corona the lab got closed, reopened and got closed again with intervals, making it hard to rely on it being open long enough so that I could do the experiment properly there. I therefore chose to do an online survey rather than a lab experiment. Such a fail testing is not possible for this experiment, but it is also not necessary. The task itself ought to be simple enough and concerning the interpretation, the sloppier the better. Highly attentive participants who analyse each sentence as if it were a puzzle are not representative of any real world sentence comprehension. They are therefore also encouraged to use their intuition when confronted with the sentence and to chose the option they *feel* is right and not what they fully *know* is right. Spending a long time on each sentence was highly

discouraged. Reaction times would be ideal in such an experiment, but due to the aforementioned shortcomings it could not be achieved.

Ideally, a reaction time would have cancelled slow answers or possibly be a variable in itself. It has after all been previously discussed that differences in reaction times alter the interpretations of the sentence. Namely that people tend to accept the sentence in a much faster pace than when declining it (Christensen, 2016, 142). It would be interesting to see which options would appear as most attractive for fast and slow responses and if there are any correlations with Christensen's data. This will for now rather be a suggestion for future research.

As of now, the only thing I can rely on is that the participants follow the instructions that I have given, which is less than ideal. Nonetheless, given the fact that the participants were all students who were all eager to help me with the project. I will trust that they have not neglected the instructions I have given them.

The sentences were all presented in the same order for all the participants. This gives me a chance to evaluate how participants gradually come to interpret the sentences as well. Whether they interpret the sentences differently in the beginning and towards the end of the survey. What such results could indicate is a learning effect on the sentences.

After every participant had filled in the survey, the data was exported as an excel sheet. One of the bigger problem with Machmaker is the problem of conditioning the sentences, meaning the sentences in the experiment themselves were not tagged. However, this was easily solved due to the similar phrasing of the different options. For instance, every option which supported an *only me* interpretation all had similar phrasings like:

(37) It was not only the X who was Y.

The same applies to the other answers as well. Any phrase with *has fewer* would most likely be the CD answer, *There's a larger amount* the CE answer and *was more often* generally attributed to SC. The data was thus easily re-tagged ready to be evaluated.

### 3.1.5 Participants

Some considerations were made for who could take the experiment. Namely that they were all Norwegian native speakers and students in their 20's with-

out any reading difficulties i.e. dyslexia.

The reason for these restrictions are, first of all, Norwegian L1 speakers are the target of interest in this study. Non-native speakers may have an interference from their native language which will affect the interpretation of the Norwegian sentences that are presented throughout the experiment.

Secondly, students tend to be more attentive than none students. If students will be tricked by the illusion, it is likely that the general population will as well.

Third, I wanted to keep the group of participant as homogenous as possible. If more age groups were added, sociolinguistic parameters would have to apply as well, which is not the interest taken in this study. Age was therefore restricted. What the experiment itself tries to figure out has to do with sentence understanding. In the experiment, sentences are presented in written form. This means someone with reading difficulties like dyslexia is not representative for the general population by means of interpretation. If the sentences were presented orally, this parameter would not have been applied.

The participants were gathered using the snowball effect. I first sent the link to the experiment to friends and acquaintances who were all university students. I encouraged many of them to further send it to other friends from other universities as well. This gave me a wide range of answers from all kinds of students from several universities across Norway.

With an experiment, presented the way I have illustrated, it is harder to account for who takes the experiment. In a lab, I have full control over who the participants are, and I can make sure they fit the group I am looking for. This is not as easy when it is all on the internet. Some of the participants, who I was not aware had dyslexia, told me that they could not take the survey for the following reason. Such honesty has been highly appreciated and hopefully there are no exceptions to this case.

## **3.2 Experimental paradigm**

Some experimental needs had to be sacrificed. The original plan was to conduct the experiment in the data lab with reaction times. The following conduction of the experiment would control the environment much more than the survey is capable of. For instance, I have previously discussed the linguistic intuitions. How do participant understand a sentence if immediately presented. They are given a short time to put it together, rather than a setting where they will

be given time to potentially analyse the content. Reaction times would have been a good way to distinguish such parameters, unfortunately that was not possible due to corona. It is therefore something I will rather suggest for a follow up study, where we could properly distinguish those parameters.

It should therefore be mentioned that the use of a survey is therefore a choice made by caution and partly by ethical concern (see next section) and not because I consider this method superior in any measure.

The data was first analysed using preliminary association tables, where each option was measured from a standard distribution (25%). No extra parameters was used in the experiment, like different sentence conditions i.e.  $\pm$ *repeatability*. This is unfortunate, given the scope could've been a little bit bigger and provided more data points for the analysis. Such parameters would provide different means of experimenting with the data. However, as the study only holds one data point such methods had to do. The grouping, as I have previously stated, was very homogenous so that general cognitive capabilities will be the main focus rather than sociolinguistics. This does not mean, however, that I do not consider sociolinguistics not fruitful. Rather, the case being I rather wanted to see how some in general tend to analyse the illusion and then leave sociolinguistic parameters for future studies.

### 3.3 Ethics

Some considerations were made for practical and ethical reasons during the conduction of the experiment. To do experiments during Covid proved itself to be a challenge on a practical level. The lab had its pockets when it was open, but it was never guaranteed when it would close, or when it would open again for that matter. This is one of the reasons why I decided to leave the lab and collect data by means of a survey instead. The practical and the ethical aspects in this situation are one and the same. If we as scientists want to protect the subjects during our experiments, then it would not be advisable to ask them to come to a lab, where potentially a lot of other people could be as well and as a consequence escalate the pandemic.

Consent was given by clicking the submit button at the end of the survey. By consenting, they agree that the data will be used in the analysis. However, this does not mean that any of their personal will be collected or saved in any capacity. It is assured in the beginning of the experiment that it is completely anonymous. It is to be said, that the experiment is tricky, and I

personally believe people will choose statements that are not complementary to the sentence presented. Such mistakes should not be scolded and should not be reflective of the competence of the participants (here used in the colloquial sense). This is, after all, not an IQ test. It is possible the survey will remind some people of language tests they did back in grad-school, and may induce an unnecessary amount of stress in people. They are therefore reminded in the text explaining the task they are about to perform that I am after the interpretation which seems the most correct to them and not according to some grammatical standard. Their data will be evaluated for scientific purposes, not for grading them.

Their results will be interesting, no matter how they perform. There are, in that sense, no real wrong or right answers, only interesting findings. I have not been interested in bettering the language of the participants, but understanding how their minds work. The goal of this experiment is not to judge their skills in understanding sentence, but of a more general cognitive competence, which may be a general feature of how humans understand sentences, rather than competence in the colloquial sense, which could be considered more akin to IQ. In the data section I will discuss some individual responses, but call them by their name. Who any of the participants are is as unknown to me as they will be to everyone else. Since the participants are kept anonymous this way, I will defend the discussion of individuals in the experiment, since it is not known who they are or no means of tracking exactly who it is.

## 4 Data and analysis

### 4.1 Data

In this section the data will be examined and then eventually analysed. I have already discussed in the section *Experimental paradigm* the statistical methods applied, I will therefore not discuss the statistical methods in any exhaustive manner here. The data will be broken down in different ways. The results cannot be understood only from the total sum, but should also be considered individually. This is why the data will be handled across participants and test items. Some of the replies may vary a lot between each other. Some participants may have very different answers from the general population, and likewise; some sentences have a very high or low response where the general reception runs contrary. In these cases, an analysis of the quality of the sentence or answering patterns of a participant will have to be analysed and explained as a deviation from the norm.

#### 4.1.1 Can the hypotheses be supported?

The  $H_0$  stated that there are no difference in variation between the different interpretations in question. Before I discuss the null hypothesis I would like to discuss the expected values contrary to the observed values. In the table below you see the observed frequencies compared to the expected frequencies.

Sentence	CD	CE	OM	SC	Sum
Observed	162	397	183	658	1400
Expected	350	350	350	350	1400
Difference	188	47	167	308	0

Table 8: The observed and expected values of each interpretation and the differences between them

As is apparent in table 8, some of the observed frequencies deviate from the expected frequencies. Some of them quite severely.

SC is by far the most frequent interpretation, indicating that it is at least a



popular interpretation. I will discuss reasons as to why in the analysis section. The most under represented interpretations are CD and OM and CE seems to be pretty close to the expected value. Below you can see the calculations made in R where the equation looks the following:  $(\text{observed-expected})^2/\text{expected}$  for each interpretation, which gives us the total sum of 458,02.

$$188*188/350 + 47*47/350 + 167*167/350 + (308*308)/350 = 458.02$$

And here are the  $\chi$ -square results from the overall results:

$$\chi\text{-squared} = 458.02, \text{df} = 3, \text{p-value} = 0.022\text{e-}14 \text{ ***}$$

As we can see the  $H_0$  can be firmly denied. There are some major variances in frequency between the different interpretations. Among the frequencies of interpretations, SC is by far the most interesting one due to the high frequency and overall deviation from expected value.

The largest contribution is SC  $(308*308)/350 = 271$  (59.2% of the sum squared value i.e.  $271/458$ ). This indicate that it is by far a very potent interpretation. This outcome will be further discussed in the analysis section. If the  $H_0$  turned out to be true, it would also be assumed the participants would make a coherent interpretation of comparative illusions. The answers would be completely random and therefore not have any effect one way or the other. The current results indicate that we interpret the sentence by means of semantic coercion, which is pretty interesting when we take into consideration that it is not an elliptic solution and doesn't meet the grammatical standards to mean what the structure entails.

My second hypothesis was that there would be a learning effect The participants will choose the non-elliptic options early in the survey, but then choose elliptic ones. This is believed due to a learning effect that can occur when the participants are exposed to the sentences enough times. The elliptic sentences are, as I have previously discussed, grammatical in some capacity and should therefore be the most attractive parsing after a learning effect. before the learning effect I believe *the good enough approach*

There are also some very unexpected responses from some individuals and

some specific sentences where some replies are a lot more frequent than what they otherwise should be. I will therefore go through some tables where the individual responses and the individual sentences are displayed and I will give reasons as to why some participants have favoured some interpretations over others.

#### 4.1.2 Individual responses

In figure 6 each participants choice is represented as a percentage. As is apparent in this table, the overall SC interpretation seems to be pretty popular; however, this is not universal, since CE is also very popular among some of the participants as well. This difference in popularity in interpretation suggests that the parsing is not universal and that different participants will in fact parse differently, possibly based on parameters not present in the current experiment. They can be based on sociolinguistic parameters, or that people emphasize different aspects of the sentence when interpreting it. On the next two pages you will see a table of every participant and their interpretation of the sentences in general and the percentage every interpretation was chosen. The same data is also shown again in figure 8, where it might be more apparent which answers stick out. This was done with Pearson Residual, which is a

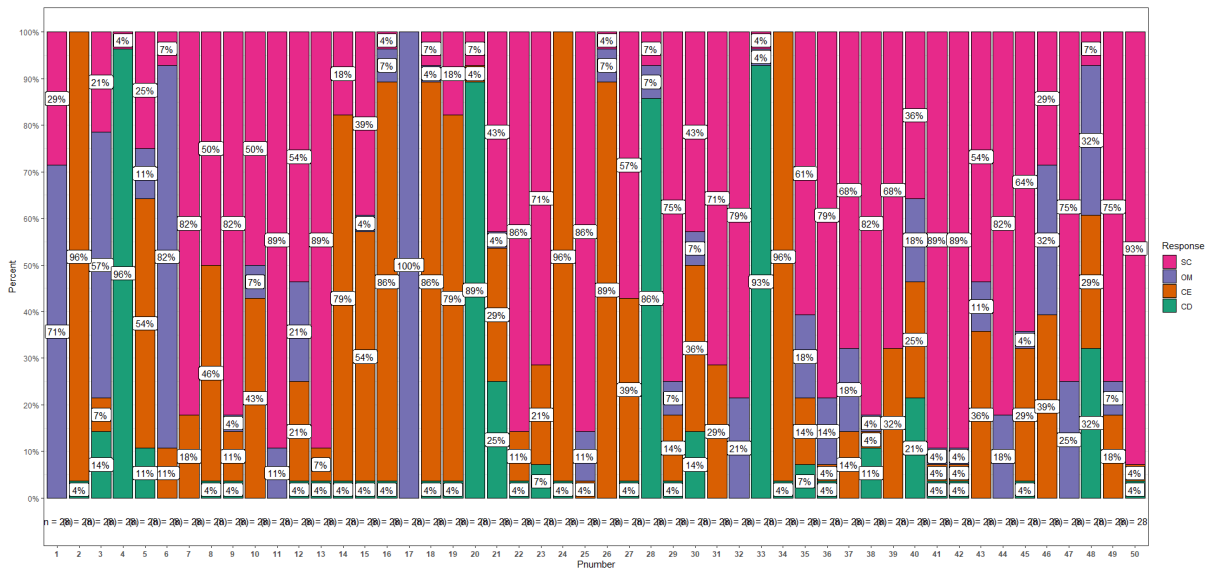


Figure 6: Participants and percentage

chi-square test for independence. What figure 6 represents is the variance from the norm. If any of the graphs is at point 0 it meets an average frequency. If above or below it is either over or under the average.

In the graph the responses from each participant in the project is enumerated from 1-50 horizontally and the different responses (CD, CE, OM and SC) vertically.

There's a big variety in the frequency of replies between the participants. Different participants are inclined towards different interpretations of the sentence, which means there is not one way the sentences are interpreted. In table 6 you will see an overview of all the participants, the amount of times each interpretation was chosen and how large a percentage it represent for the total amount of interpretations chosen.

As is apparent in figure 6, where the different interpretations are measured up to one another using Pearson Residual, is that there are some responses that stick out like a sore thumb. I will go through these to explain why some of these stick out as much as they do. The one interpretation that seems to consistently stay high is SC. The low frequency of SC seems to be paired with a high frequency of another interpretation instead. It should therefore be concluded that SC is more or less the standard interpretation of the sentence and that there are variables not accounted for which makes other interpretations as strong in some of the sentence. The interpretations that seem to take over for SC in some sentences are very frequent where they occur, but are otherwise low. It is therefore most likely something to do with the sentences in particular and not a general parsing of the sentence.

CD and OM are generally very low in frequency, but are very high among some participants. For instance; participant 18 chose OM for all of the sentences, even though OM is generally not a popular alternative for most participants. The reason for this could be that this participant went through the experiment in a hasty manner. They maybe found a pattern from the first comparative illusion presented and picked the OM option based on the evaluation made from the first sentence. It is unfortunately not possible to check whether participants have been thorough or not, but I will come back to some means to know that other participants generally seem to have been thorough when picking interpretations. In regards to outliers, there are others who have found otherwise unpopular interpretations to be very attractive and picked them at a very high frequency. This is true for participant 4, 20, 28 and 33

.id	label	variable	Response			
			CD	CE	OM	SC
		P10	1 (3.57%)	3 (10.71%)	1 (3.57%)	23 (82.14%)
		P11	0 (0%)	12 (42.86%)	2 (7.14%)	14 (50.00%)
		P12	0 (0%)	0 (0%)	3 (10.71%)	25 (89.29%)
		P13	1 (3.57%)	6 (21.43%)	6 (21.43%)	15 (53.57%)
		P14	1 (3.57%)	2 (7.14%)	0 (0%)	25 (89.29%)
		P15	1 (3.57%)	22 (78.57%)	0 (0%)	5 (17.86%)
		P16	1 (3.57%)	15 (53.57%)	1 (3.57%)	11 (39.29%)
		P17	1 (3.57%)	24 (85.71%)	2 (7.14%)	1 (3.57%)
		P18	0 (0%)	0 (0%)	28 (100.00%)	0 (0%)
		P19	1 (3.57%)	24 (85.71%)	1 (3.57%)	2 (7.14%)
		P2	0 (0%)	0 (0%)	20 (71.43%)	8 (28.57%)
		P20	1 (3.57%)	22 (78.57%)	0 (0%)	5 (17.86%)
		P21	25 (89.29%)	1 (3.57%)	0 (0%)	2 (7.14%)
		P22	7 (25.00%)	8 (28.57%)	1 (3.57%)	12 (42.86%)
		P23	1 (3.57%)	3 (10.71%)	0 (0%)	24 (85.71%)
		P24	2 (7.14%)	6 (21.43%)	0 (0%)	20 (71.43%)
		P25	1 (3.57%)	27 (96.43%)	0 (0%)	0 (0%)
		P26	0 (0%)	1 (3.57%)	3 (10.71%)	24 (85.71%)
		P27	0 (0%)	25 (89.29%)	2 (7.14%)	1 (3.57%)
		P28	1 (3.57%)	11 (39.29%)	0 (0%)	16 (57.14%)
		P29	24 (85.71%)	0 (0%)	2 (7.14%)	2 (7.14%)
		P3	1 (3.57%)	27 (96.43%)	0 (0%)	0 (0%)
		P30	1 (3.57%)	4 (14.29%)	2 (7.14%)	21 (75.00%)
		P31	4 (14.29%)	10 (35.71%)	2 (7.14%)	12 (42.86%)

P	P	P32	0 (0%)	8 (28.57%)	0 (0%)	20 (71.43%)
		P33	0 (0%)	0 (0%)	6 (21.43%)	22 (78.57%)
		P34	26 (92.86%)	0 (0%)	1 (3.57%)	1 (3.57%)
		P4	4 (14.29%)	2 (7.14%)	16 (57.14%)	6 (21.43%)
		P40	0 (0%)	9 (32.14%)	0 (0%)	19 (67.86%)
		P41	6 (21.43%)	7 (25.00%)	5 (17.86%)	10 (35.71%)
		P42	1 (3.57%)	1 (3.57%)	1 (3.57%)	25 (89.29%)
		P43	1 (3.57%)	1 (3.57%)	1 (3.57%)	25 (89.29%)
		P44	0 (0%)	10 (35.71%)	3 (10.71%)	15 (53.57%)
		P45	0 (0%)	0 (0%)	5 (17.86%)	23 (82.14%)
		P46	1 (3.57%)	8 (28.57%)	1 (3.57%)	18 (64.29%)
		P47	0 (0%)	11 (39.29%)	9 (32.14%)	8 (28.57%)
		P48	0 (0%)	0 (0%)	7 (25.00%)	21 (75.00%)
		P49	9 (32.14%)	8 (28.57%)	9 (32.14%)	2 (7.14%)
		P5	27 (96.43%)	0 (0%)	0 (0%)	1 (3.57%)
		P50	0 (0%)	5 (17.86%)	2 (7.14%)	21 (75.00%)
		P51	1 (3.57%)	1 (3.57%)	0 (0%)	26 (92.86%)
		P6	3 (10.71%)	15 (53.57%)	3 (10.71%)	7 (25.00%)
		P7	0 (0%)	3 (10.71%)	23 (82.14%)	2 (7.14%)
		P8	0 (0%)	5 (17.86%)	0 (0%)	23 (82.14%)
		P9	1 (3.57%)	13 (46.43%)	0 (0%)	14 (50.00%)

Figure 7: The frequency of each response per participants including percentage

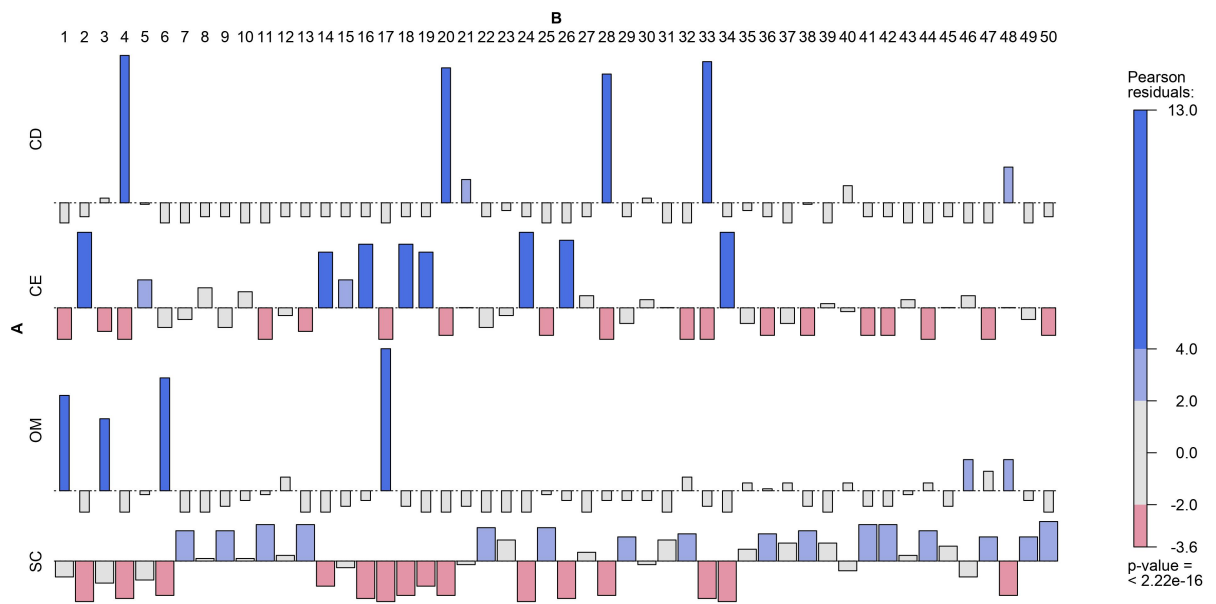


Figure 8: General responses per participant (1-50) to the comparative illusions and the frequency of responses. The bar is the average amount of responses and the blue and the red represent significant derivations from the norm

who all found CD to be a very attractive interpretation of the comparative illusion.

#### 4.1.3 Response per sentence

As is apparent in figure 8, where the different interpretations are measured up to one another using Pearson Residual is that there are some responses that stick out like a sore thumb. I will go through these to explain why some of these stick out as much as they do. The one interpretations that seems to consistently stay high is SC. The low frequency of SC seems to be paired with a high frequency of another interpretation instead.

It should therefore be concluded that SC is more or less the standard interpretation of the sentence and that there are variables not accounted for which makes other interpretations as strong in some of the sentence. The interpretations that seem to take over for SC in some sentences are very frequent where

they occur, but are otherwise low. It is therefore most likely something to do with the sentences in particular and not a general parsing of the sentence. Figure 8 shows the general response to each of the sentence they were exposed to. As we see, there are some sentences with a large over-representation of CD, CE or OM. The lack of over-representation of SC may be due to the overall high frequency of the interpretation. None of the sentences are conditioned, meaning there shouldn't be that high over-representations in the data. I will proceed to explain the individual examples put forward as good as possible. Sentence 8 is the following one:

- (38) Flere bjørner har gått i hi for vinteren enn grevlingen har. (S8)  
 more bear-pl has gone in place-of-hibernation for winter-det than  
 badger-det has(aux/pos)  
 ”More bears have been hibernating in its cave more than the badger  
 has”

This sentence is in no way particularly special. It is not repetitive, nor does it seem to have any general grammatical quirks to it which would make the sentence stand out.

Why CD is overrated for this particular sentence has nothing to do with the sentence itself, but rather one of the possible replies. In the SC solution, which

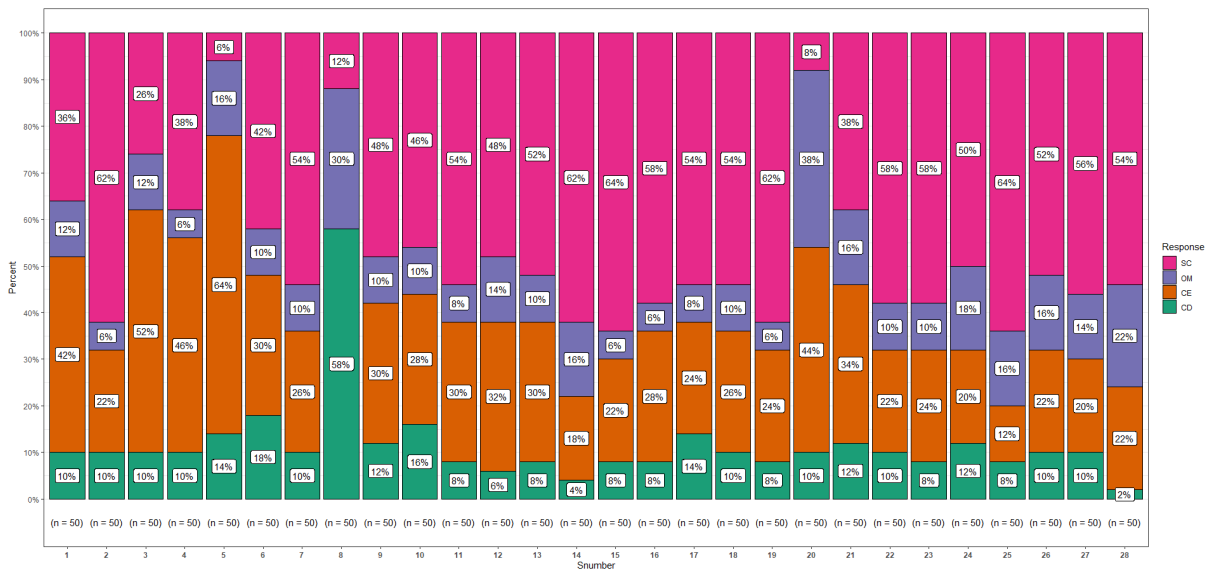


Figure 9: The percentage of each sentence in the experiment

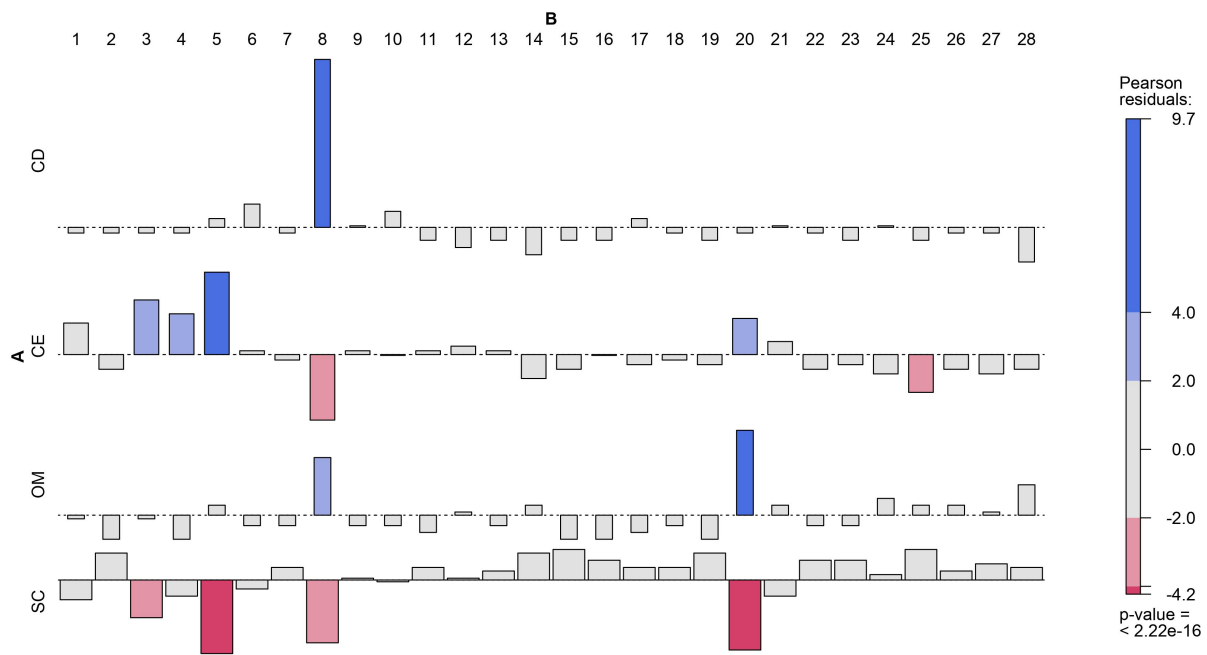


Figure 10: The general response per sentence. The bar is the average amount of responses and the blue and the red represent significant derivations from the norm.



is a solution who otherwise seem to be pretty popular, missed a word. This is probably why SC is so unpopular in this sentence to begin with as well. In the SC sentence a *more* was missing, which makes the sentence ungrammatical and therefore less

- (39) \*Bjørnene gikk i hi enn grevlingen har. (S8, SC)  
bear-pl.det went in place-of-hibernation than badger-det has(aux/pos)  
\*”The bears went into hibernation for the winter than the badger has”

What this sentence was supposed to say is *the bears went into hibernation more times than the badger has*. This error has some perks to it. First of all, it is proven the overall popularity of SC. It seems after all that it is the mistake that causes the over-representation of CD. This error also suggests CD to generally be a good second choice for many people, when SC fails them. It would’ve been interesting to see in a future experiment whether this would be the case if participants were to also grade the second best sentence as well. Which is an interesting suggestion, given that CD is not very popular if only the most attractive parsing is considered.

Why CD? One of the most likely reasons is that the sentence can be read as non-repetitive, meaning CD would be a more likely contender. CE, which is in general a more popular contender when SC is low is in this scenario particularly low. Whether the sentence can be read as non-repetitive depends how *for the winter* is read. Is it the same winter or could it be several? If several winters then *for the winter* is only a specifier, meaning the if the bear or badger enters the place of hibernation for a night it doesn’t count. Meaning they both have entered their respective places for hibernation with the intention of sleeping through the winter. The amount would then rely on how many winters the bear has lived compared to the badger.

If we read *the winter* as only the current one, then we don’t have a repeating event, since the bear(s) and the badger can only go into hibernation for the winter once. It is possible this ambiguity has left CE less attractive.

Finally, it proves that most of the participants paid close attention to the sentences individually during the experiment. It is reason to believe that if the interpretation with an error had been picked or if the error had not made any significant differences in choice, that we could likewise say that the participants had been sloppy with the handling of the data or that a *good enough approach* could be applied to explain the phenomenon. However, this is not the case. They chose away SC for a minor error, meaning they are conscientious enough

to notice the error and chose SC away in favour of CD, but not enough to see the fault in SC in general. This will be further explored in the analysis section.

In sentence number 5 SC is low and CE has taken its place as the most frequent interpretation. CE was chosen 32 out of 50 times, and the overall frequent SC only 3 times. This makes for a big difference in frequency, where it seems as if CE has taken the place of SC.

- (40) Flere elger har løpt gjennom skogen enn jerven har. (S5)  
more moose-det.pl AUX run.past through forest-det than wolverine-  
det.sg has(aux/pos)  
*More moose have been running the forest than the wolverine has*

What seems to be the case is that people do not believe the illusion. They understand that there are more moose and not more amount of times someone has been running through the forest. Why do they not get tricked by the illusion in this sentence? If there is any particular variable that makes CE the most attractive parsing in 5, sentence 3 and 4 may have the same feature, given that CE are very high in those sentences as well.

- (41) a) Flere harer har hoppet over gjerdet enn geita har. (S3)  
more hare-det.sg AUX jump.pst over fence-det.sg than goat-det.sg has(aux/pos)  
*More hares has jumped over the fence than the goat has*
- b) Flere elger har krysset bekken enn ulven har. (S4)  
more moose-det.pl AUX cross.pst pond-det.sg has(aux/pos)  
*More moose has crossed the pond than the wolf has*

3,4 and 5 have some things in common. I will list them and say how likely it is which feature it is that gives all of them such a low score of SC and a high score in CE respectively. As I have previously mentioned, CE is grammatically sound due to the resolution of the ellipsis in the original sentence. This is not the case with SC, where an adjacent meaning is applied. It is therefore a possibility that these sentences are not as illusory as the rest.

What sentence number 3,4 and 5 has in common is that they have a noun the action in question is associated with. A fence, a pond and a forest. These are all singular, meaning they cannot be misleadingly quantified. The suggestion is that people in general quantify the amount of fences ponds and forests,

meaning i.e. the hares has jumped over more fences, rather than jumped over the same fence more times.

SC should not in theory be as low, considering in the interpretation, the amount is the time someone has jumped over the fence, crossed the pond or been running through the forest, not the amount of fences, ponds and forests, and should therefore not play any pivotal role. The fact that this feature is common across so many sentences with a low SC is suspicious. However, there are more sentences out there with such a feature, where SC is around normal frequency, like 28.

- (42) Flere kaniner har knasket på gulrota enn haren har (S28)  
more rabbit-det.pl AUX munched on carrot-det.sg than hare.det.sg  
has(aux/pos)  
*More rabbits has munched on the carrot than the hare has.*

In this sentence it's only one carrot, SC was chosen 54% of the time, meaning they did not interpret this sentence to mean the rabbit ate more carrots, but rather that it is only one carrot and the rabbit ate more of it than the hare. The verb itself and the following preposition could be more suggestive of there being just one carrot. In the diaglossia I have translated *knaske* into *munch*, which is indeed the closest word to translate it, however, it is not a perfect translation. *Nib* could be a worthy translation as well, in the sense that nib is suggestive of the fact that not the entire carrot was eaten and that we therefore can measure the quantity of one carrot being eaten and not the amount of carrots, which would otherwise be the most normal way of measure. It is however not adequate, given the manner of eating is different. In the sentence it could hold some associations that the rabbit would eat carrots, yet the hare wouldn't. Given these traits, it is possible for 28 to be an exception and for the singular nouns in 3,4 and 5 to be the reason for the low frequency of SC.

This, however, does not explain why CE is complementarily high. Below you can see the CE and SC options for 3,4 and 5. I will first discuss them individually, look for oddities that have been overlooked initially and could have affected the overall results and then compare them to sentences where the SC is high and the CE is on an average low. If there are some distinct differences between these traits, we can conclude why SC is low or why CE is as high as it is.

- (43) a) Det er et større antall harer enn ganger geita hoppet  
over gjerdet (3, CE)

It is a.neutr big-sup amount hare-det.pl than times goat-det.sg jumped over fence-det.sg

*It is a bigger amount of hares than the amount of times the goat jumped the fence*

b) Haren hoppet over gjerdet flere ganger enn geita (3, SC)

Hare-det.sg jumped over fence-det.sg more times than goat-det.sg

*The hare jumped over the fence more times than the goat*

c) Det er et større antall elger enn ganger ulven krysset bekken (4, CE)

It is a.neutr big-sup amount moose-det.pl than times wolf-det.sg cross.PST pond-det.sg

It is a bigger amount of mooses than times the wolf crossed the pond.

d) Elgen hoppet over bekken flere ganger enn ulven (4, SC)

moose-det jump.prst over pond more times than wolf-det

The moose jumped over the pond more times than the wolf

e) Det er et større antall mus enn ganger rotta var i åkeren. (5, CE)

DET be.prst big.COMP amount mouse.pl than times rat.det was in field-det

It is a bigger amount of mice than times the rat was in the field

f) Musa var oftere i åkeren enn rotta. (5, SC)

Mouse.det.sg be.pst often-COMP in field-DET than rat-DET

The mouse was more often in the field than the rat

One possible explanation for the overall high frequency of CE in 3,4 and 5 could be the fact that they occur early in the experiment. Sentence 1 is not significant, but higher than average, sentence 2 is the exception, which is lower. It could therefore be that CE is to some extent the interpretation many people cling to initially, but will eventually turn away from. Despite its grammaticality, the meaning might be a little too strange for most people to make it acceptable. The overall low frequency of SC, which is very apparent in 3 and 5 might be a better explanation, it is also quite low in 4, but not to a significant extent.

In S20 has an over than average representation of OM, but has the highest

frequency of CE.

- (44) Flere rever har besøkt bondegården enn grevlingen har. (S20)  
more fox-pl has visited farm-pl than badger-det.sg has(aux/pos)  
More foxes has visited the farm than the badger has

This sentence shares the singular noun connected to the action, like 3,4 and 5. This might indicate that an object connected to the verb somehow makes for a high frequency of CE. Why OM occur can first of all be because it appears at the top among the interpretations. It might also has something to be with the pragmatics. We will often associate the fox with going to a farm, which occurs in many songs and stories(at least in Norway). To then receive the sentence that it was not only the badger might give the participants some associations which might have affected how they have interpreted this sentence.

- (45) Det er flere enn bare grevlingen som besøkte bondegården (S20, OM)  
it is more than just/only badger-det.sg which visit.pst farm-det.sg  
*It's not only the badger who visited the farm*

A general tendency of the OM interpretations is that they are generally true, but not very relevant to the sentence. It is true as a general presupposition that there are are more than the badger who has visited the farm, given the fact that it says more foxes has visited the place. People generally don't seem to choose this interpretation, despite its technical correctness. Participant 18 is an exception, I have discussed the choices of this participant in the previous section. People seem to choose the sentences that are relevant, rather than presupposed. But that doesn't explain why it is particularly here in this case. Despite the technical correctness of OM, it was usually chosen away for other interpretations.

#### 4.1.4 Second hypothesis: Learning effect

You will here see two tables that represent the overall numbers per sentence in the experiment. My second hypothesis was that a learning effect would during the experiment, especially since the fact that some options like SC may look attractive on the surface, since it's a much simpler interpretation than any of the elliptic adequate ones, but that the participants would eventually figure out the trick with these sentences the more they would be exposed to them. As we will see with the figures below this was not the case.

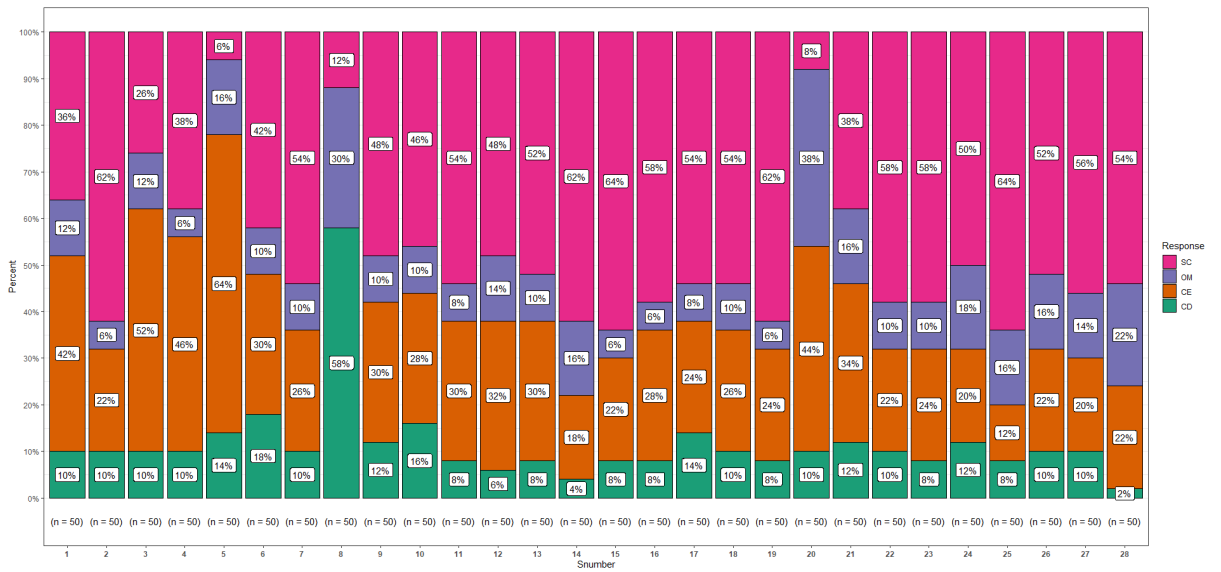


Figure 11: Sentences and percentage per sentence. From left to right you see the sentences in the same order as they were presented in the experiment. As should be apparent, there are no major variations from start to finish, which indicate that my second hypothesis about a learning effect can be rejected.

.id	label	variable	Response			
			CD	CE	OM	SC
		S1	5 (10.00%)	21 (42.00%)	6 (12.00%)	18 (36.00%)
		S10	8 (16.00%)	14 (28.00%)	5 (10.00%)	23 (46.00%)
		S11	4 (8.00%)	15 (30.00%)	4 (8.00%)	27 (54.00%)
		S12	3 (6.00%)	16 (32.00%)	7 (14.00%)	24 (48.00%)
		S13	4 (8.00%)	15 (30.00%)	5 (10.00%)	26 (52.00%)
		S14	2 (4.00%)	9 (18.00%)	8 (16.00%)	31 (62.00%)
		S15	4 (8.00%)	11 (22.00%)	3 (6.00%)	32 (64.00%)
		S16	4 (8.00%)	14 (28.00%)	3 (6.00%)	29 (58.00%)
		S17	7 (14.00%)	12 (24.00%)	4 (8.00%)	27 (54.00%)
		S18	5 (10.00%)	13 (26.00%)	5 (10.00%)	27 (54.00%)
		S19	4 (8.00%)	12 (24.00%)	3 (6.00%)	31 (62.00%)
		S2	5 (10.00%)	11 (22.00%)	3 (6.00%)	31 (62.00%)
		S20	5 (10.00%)	22 (44.00%)	19 (38.00%)	4 (8.00%)
S	S	S21	6 (12.00%)	17 (34.00%)	8 (16.00%)	19 (38.00%)
		S22	5 (10.00%)	11 (22.00%)	5 (10.00%)	29 (58.00%)
		S23	4 (8.00%)	12 (24.00%)	5 (10.00%)	29 (58.00%)
		S24	6 (12.00%)	10 (20.00%)	9 (18.00%)	25 (50.00%)
		S25	4 (8.00%)	6 (12.00%)	8 (16.00%)	32 (64.00%)
		S26	5 (10.00%)	11 (22.00%)	8 (16.00%)	26 (52.00%)
		S27	5 (10.00%)	10 (20.00%)	7 (14.00%)	28 (56.00%)
		S28	1 (2.00%)	11 (22.00%)	11 (22.00%)	27 (54.00%)
		S3	5 (10.00%)	26 (52.00%)	6 (12.00%)	13 (26.00%)
		S4	5 (10.00%)	23 (46.00%)	3 (6.00%)	19 (38.00%)
		S5	7 (14.00%)	32 (64.00%)	8 (16.00%)	3 (6.00%)
		S6	9 (18.00%)	15 (30.00%)	5 (10.00%)	21 (42.00%)
		S7	5 (10.00%)	13 (26.00%)	5 (10.00%)	27 (54.00%)
		S8	29 (58.00%)	0 (0%)	15 (30.00%)	6 (12.00%)
		S9	6 (12.00%)	15 (30.00%)	5 (10.00%)	24 (48.00%)

Figure 12: The comparative illusion or filler sentence presented on top and four options below each one. Each option represents a different interpretation of the sentence above.

As a matter of fact, the effect seems almost to be the opposite as the participants seem to pick SC more often the further into the experiment they get. CD and CE seem to go from being mildly popular in the beginning to very unattractive in the later stages. It is therefore safe to say that the elliptic interpretations are not generally detected by people, might be so by some individuals, but it is not the norm.



## 4.2 Analysis

### 4.2.1 What can be said about the results?

In section 4.1.1 I displayed a general frequency of the interpretations and a  $\chi$ -square table which was very significant, meaning the null hypothesis can be rejected. The most frequent alternative was SC which overall made up 59% of the sum of the squared value. The high frequency of SC was expected, due to similar conclusions from previous research.

In section 1.4.2 I went through the results from Wellwood et al. (2018), where several variants of the sentence were used which consequently also cancelled out certain interpretations of the sentence. What she was left with was the possibility of event comparison being the only viable option for interpretation (called SC throughout this paper). Despite the fact that the interpretation doesn't cohere to any elliptic solution, it is nonetheless the most attractive parsing. It means that there are some interesting things going on when we interpret the sentence, which is independent of the grammar of the sentence. I will highlight some of the theories I have previously discussed in chapter 1, which seem to have some explanatory power as to why the participants understand CI to mean something outside of the grammatical constraints.

I will not do so exhaustingly, because they have already been discussed throughout the aforementioned chapter, but will be reiterated for the sake of clarity. The fact that the experiment was conducted in Norwegian with L1 speakers should bear some consideration, regardless of what I have discussed in 1.2.2, that there are very small syntactic and semantic differences between the sentence in Norwegian and English. Norwegian and Danish are even more similar due to historical reasons and due to the fact that both languages belong to the Northern Germanic language family. In 1.7.1 I discussed a survey by Christensen (2016) where the sentence was interpreted by Swedish and Danish students in Danish. There are some Danish idiosyncrasies, like example a) from Table 1, reiterated below as 46a) It has already been discussed in 1.7.1 why this only occurs in Danish and therefore not applicable to Norwegian and will therefore not be further discussed in this section.

- (46) a) ... [Except me]  
b) ... than just me  
c) ...[more (often) than I have]  
d) It doesn't make sense

e) ... [Than I own]

The reason why this example is the most frequent one, and why it does not matter in the current Norwegian experiment is discussed in 1.7.1 and therefore not further discussed here. The second highest was sentence b), which correlates with the OM interpretation, which was far below the expected value in the current experiment. Reasons for this difference could either be due to the method applied in the current experiment, the language differences or the fact that the students handed in a paper on what they thought the sentence meant, contrary to the hasty decisions the current experiment encourages. There does not seem to be much ground for assuming any linguistic variables to be the determining factor in this sentence. The sentence reconstructing the meaning outlined in 47.

(47) It was not only the X who Y

It may not be as attractive a rendering as other ways to come to the same conclusion. Time could be a factor to play in as well. The Swedish and Danish students had plenty of it to truly consider the content of the sentence, which is not the case for most other experiments on the same topic. Albeit, it is not likely that it's what made the difference, due to the fact that it is not a grammatically coherent option and therefore shouldn't be a conclusion people come to solely based on the fact that they have considered for a longer time. My method therefore may have a saying in the interpretation of the sentence. Option C) aligns with SC, which was the least frequent in the Swedish/Danish experiment. SC was the most frequent in my experiment and has showed itself to be an attractive parsing in English as well Wellwood et al. (2018). The low frequency in Christensen's experiment marks itself as an anomaly compared to the rest of the research on interpretation. The most likely contender to this anomaly seems to be time. The interpretation is not grammatically sound and should be rejected through tough thought. The next option will not be discussed, since it is not permitted in the present study.

The final option presented, 46e) aligns with the CD interpretation. The option is grammatically sound, as has also been acknowledged by Christensen. It is not very frequent in any of the experiments, which suggests that it might not be an attractive option overall.

The fact that CE was not even mentioned in the Christensen survey hints that it's not even considered an option when parsed actively (without options). CE was generally high in the present experiment, but it could likely be because it

was presented to the participants who might have recognized the sentence as coherent. The linguistic difference could be a factor here as well, option a) is an explicitly Danish option, which might cancel out several other options due to the attractiveness of aforementioned interpretation.

#### **4.2.2 The ungrammatical interpretation**

In general SC is a very attractive parsing of the sentence and statistically significant in its difference in frequency. This is despite the fact that it is not a grammatical coherent interpretation. It is after all not possible that the sentence depicts a comparison between events as I have discussed in section 1.4.2. I have also previously made some hints as to why such a sentence can be as attractive as it is, despite this fact.

In section 1.8.1 I discussed Grice (1975) and his maxims, Tversky and Kahneman (1983) where the Linda example was exhibited. In that section I discussed how these concepts complement each other and how they connect with comparative illusions by means of quick reasoning. In such scenarios the good enough approach (Ferreira et al., 2007) applies, since a temporal limitation requires a threshold. The overall frequency of SC, rather than any elliptically adequate solution suggest the replies were overall fast and judged intuitively. We did not see any learning effects on the sentences either, which suggests that the elliptic interpretations are not necessarily acceptable for the participants, as was assumed due to the fact that they are grammatically adequate. The detectability is not present, nonetheless. The occasional individual responses seems to indicate it is for some people, but that it has more to do with individual perceptiveness and might occur for participants who like to be thorough. Reaction times would be ideal to distinguish these types of responses.

The overall non-learning effect may also suggest that the participants never become aware of any viable elliptic option during the participation and that they will stay fooled throughout. Ferreira's framework has some great explanatory power in this case, if we see the participants as never consider (as evidence suggest) the sentence above the aforementioned threshold of interpretation. The stimuli of each argument in the sentence is put forth, but not properly considered in the appropriate relation to one another, similarly to how the probability of Linda's properties in Tversky and Kahneman (1983) is not properly considered due to the stimulation caused by the associations

we have to Linda, what it means to be a bank teller and what it means to be a feminist. In general terms, the stimuli we receive seems to be vital first property of perception before we put the relations (in this case the syntax) together. What comparative illusions and the Linda example seem to indicate is that we tend to become lazy when we are put to the task of understanding these relations and might rely on stereotypes or other forms of previous knowledge about relations around us to make sense of a situation.

### **4.3 Conclusion**

What this study has tried to achieve, is to get an idea whether there are any interpretation in the mind of the recipient when exposed to comparative illusions and if they interpret the sentence in any specific way upon initial exposition. It is hard to tell if people actually interpret the sentence in any way when they first encounter it. It was not the intention to answer this question in any definite manner, but rather to give an indication of it by frequency of any interpretation in particular. The latter has been achieved. The participants were by no means neutral in their responses, which the refutation of the  $H_0$  suggest. However, as I have discussed in section 4.1.2, there are some major variations between individual responses. This suggests that people favour different interpretations, most likely on parameters not addressed within the scope of this study, which has to do with idiosyncratic ways of thinking, judging and making decisions. Sociolinguistic parameters could also apply. It is nonetheless undeniable that the data has a general tendency towards an SC interpretation of the sentence.

#### **4.3.1 Suggestions for future research**

First of all, as I have discussed in this paper to some extent. Due to Covid19 the experiment itself became rather limited to what I would otherwise like to do. I think especially reaction times would've been an interesting addition to a future experiment. If people accept the sentence quickly, do they also choose some option more quickly than others. A reaction time would also make some restrictions, as they currently could spend a lot of time on each sentence and thoroughly examine it, which is not desirable.

It would have been interesting as well to see how participants would respond if they heard the sentences rather than reading it and obviously more experi-

ments in different languages, since the current pool of languages is still quite small.

As I have briefly hinted at in the previous section that there tend to be some individual variety to the different interpretations, which might invite a qualitative study of some kind, where different trait like ways of thinking, judging etc. could be aligned with different interpretations of the sentence. If given the right parameters for testability I am sure the results could be interesting.

In section 1.1.1, where I make my motivation for the current study I outlined how grammar is considered a cognitive module, yet as we see in a case like comparative illusions that our judgements fall outside of it. The separation of parser and grammar could potentially be a fruitful new framework of syntax, where experimentation on garden path sentences and grammatical illusions could lead the way. This mantle I will leave for someone else to pick up, as I hope someone eventually will.

## References

- Abboud, H. (2015). Superlab 5. Cedrus.
- Bach, M. and Poloschek, C. M. (2006). Optical illusions. *Adv Clin Neurosci Rehabil*, 6(2):20–21.
- Bever, T. and Townsend, D. (2001). Some sentences on our consciousness of sentences.
- Bresnan, J. (1975). 'Comparative Deletion and Constraints on Transformations', *Linguistic Analysis*1, 25-74. *Bresnan251Linguistic Analysis1975*.
- Carnie, A. (2021). *Syntax: A Generative Introduction*. John Wiley & Sons.
- Chomsky, N. (1957). *Syntactic Structures*. Walter de Gruyter.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*, volume 11 of *Special Technical Report (Massachusetts Institute of Technology. Research Laboratory of Electronics)*. MIT Press, Cambridge, Mass.
- Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, 3(1):1–15.

- Chomsky, N. (1981). *Lectures on Government and Binding: The Pisa Lectures*. Mouton de Gruyter, Berlin, New York, 7th ed. edition.
- Christensen, K. (2010). Syntactic reconstruction and reanalysis, semantic dead ends, and prefrontal cortex. *Brain and cognition*, 73:41–50.
- Christensen, K. (2016). The dead ends of language: The (Mis)Interpretation of a grammatical illusion. pages 129–160.
- de Dios-Flores, I. (2016). More people have presented in conferences than I have. Comparative illusions: When ungrammaticality goes unnoticed. *Glancing Backwards To Build a Future in English Studies*, page 219.
- Ferreira, F., Bailey, K. G., and Ferraro, V. (2007). Good-Enough Representations in Language Comprehension. *Current directions in psychological science : a journal of the American Psychological Society*, 11(1):11–15.
- Fults, S. and Phillips, C. (2004). The source of syntactic illusions. In *Poster Presented at the 17th Annual CUNY Conference on Human Sentence Processing*. College Park, MD.
- Gorrell, P. (1995). Syntax and parsing. *Cambridge studies in linguistics*, (76).
- Grice, H. P. (1975). Logic and Conversation. *Speech Acts*, pages 41–58.
- Grimshaw, J. and Rosen, S. T. (1990). Knowledge and obedience: The developmental status of the binding theory. *Linguistic inquiry*, 21(2):187–222.
- Halle, M. and Stevens, K. (1962). Speech recognition: A model and a program for research. *IRE Transactions on Information Theory*, 8(2):155–159.
- Harris, R. A. (1995). *The Linguistics Wars*. Oxford University Press USA, New York.
- Hofstadter, D. R. (1979). *Gödel, Escher, Bach*. Penguin Books, Harmondsworth, Middlesex, England.
- Kelley, P. (2018). *More People Understand Eschers than the Linguist Does: The Causes and Effects of Grammatical Illusions*. PhD thesis, Michigan State University, United States – Michigan.

- Kroeger, P. (2018). *Analyzing Meaning: An Introduction to Semantics and Pragmatics*. Language Science Press.
- Kroeger, P. R. (2004). *Analyzing Syntax: A Lexical-Functional Approach*. Cambridge University Press.
- Ladusaw, W. A. (1979). *Polarity Sensitivity as Inherent Scope Relations*. The University of Texas at Austin.
- Lafer-Sousa, R., Hermann, K. L., and Conway, B. R. (2015). Striking individual differences in color perception uncovered by ‘the dress’ photograph. *Current Biology*, 25(13):R545–R546.
- Leivada, E. (2020). Language Processing at Its Trickiest: Grammatical Illusions and Heuristics of Judgment. *Languages*, 5(3):29.
- Leivada, E. and Westergaard, M. (2020). Acceptable Ungrammatical Sentences, Unacceptable Grammatical Sentences, and the Role of the Cognitive Parser. *Frontiers in Psychology*, 11.
- Liberman, M. (2004). Escher sentences. <http://itre.cis.upenn.edu/~myl/language-log/archives/000862.html>.
- Montalbetti, M. (1984). After binding: On the interpretation of pronouns.
- Napoli, D. J. (1983). Comparative Ellipsis: A Phrase Structure Analysis. *Linguistic Inquiry*, 14(4):675–694.
- Nussbaum, M. C. (2017). *Subset Comparatives as Comparative Quantifiers*. Thesis, Massachusetts Institute of Technology.
- O’Connor, E. (2015). Comparative illusions at the syntax-semantics interface. *University of Southern California*.
- O’Connor, E., Pancheva, R., and Kaiser, E. (2013). Evidence for online repair of Escher sentences. In *Proceedings of Sinn Und Bedeutung*, volume 17, pages 363–380.
- Phillips, C., Wagers, M. W., and Lau, E. F. (2011). 5: Grammatical Illusions and Selective Fallibility in Real-Time Language Comprehension. *Experiments at the Interfaces*, pages 147–180.

- Popper, K. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge.
- Reinhart, T. M. (1976). *The Syntactic Domain of Anaphora*. PhD thesis, Massachusetts Institute of Technology.
- Rosseland, L., Arcidiacono, A., Sonntag, L., Johansson, C., and Larsen, T. (2021). Experimental evidence for comparative illusions. In *The Annual Psycholinguistics in Flanders Conference*, Kaiserslautern, Germany.
- Schütze, C. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Language Science Press.
- Townsend, D. J. and Bever, T. G. (2001). *Sentence Comprehension: The Integration of Habits and Rules*. MIT Press.
- Tversky, A. and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, 90(4):293.
- Vasishth, S., Brüßow, S., Lewis, R. L., and Drenhaus, H. (2008). Processing Polarity: How the Ungrammatical Intrudes on the Grammatical. *Cognitive Science*, 32(4):685–712.
- Wellwood, A., Pancheva, R., Hacquard, V., Fults, S., and Phillips, C. (2009). The role of event comparison in comparative illusions. In *Poster Presented at the 22nd Annual CUNY Conference on Human Sentence Processing*. Davis, CA.
- Wellwood, A., Pancheva, R., Hacquard, V., and Phillips, C. (2018). The Anatomy of a Comparative Illusion. *Journal of Semantics*, 35(3):543–583.



## 5 Appendix

Here is the description given to the participants and the sentences presented to them and the options. I have here separated the CI sentences from the filler sentences, they were mixed in the experiment. I have also added the condition to each sentence here, which was not visible to the participants.

Oppgave: OBS! Les oppgaveteksten nøye slik at du forstår eksperimentet tydelig nok! I denne oppgaven kommer du til å bli presentert for setninger som alle handler om dyr i naturen. Under hver setning vil du finne fire tolkninger av setningen.

Din oppgave er å velge tolkningen som virker mest korrekt for DEG. Føl på instinktet ditt hva som er den mest korrekte tolkningen og ikke bruk for mye tid på hver setning. Om du føler at du ikke har forstått setningen velger du den tolkningen som føles rett i øyeblikket og går videre til neste setning. Lykke til og tusen takk!

1. Er norsk ditt morsmål?

Ja, Nei

2. Har du lese eller skrivevansker?

Ja, Nei

Conditions:

CD= comparative deletion (A)

SC= Semantic coercion (se Wellwood; event comparison) (B)

CE= Comparative ellipsis (C)

OM = Only me (a form of semantic coercion where: I have= me) (D)

Training sentences

CI sentences:

1 Flere gauper har vært på fjellet enn hjorten har

... Hjorten har færre gauper enn de som var på fjellet. (CD)

... Gaupen var oftere på fjellet enn hjorten. (SC)

... Det er et større antall gauper enn ganger hjorten var på fjellet (CE)  
... Det er ikke bare hjorten som var på fjellet (OM)

2 Flere elger har brukt treet til å kvesse giviret sitt enn hjorten har  
... Det er et større antall elger enn ganger hjorten brukte treet (CE)  
... Det er flere enn bare hjorten som brukte treet (OM)  
... Elgene brukte treet mer enn hjorten (SC)  
... Hjorten har færre elger enn de som brukte treet (CD)

3 Flere elger har krysset bekken enn ulven har.  
... Ulven har færre elger enn de som krysset bekken (CD)  
... Elgen hoppet over bekken flere ganger enn ulven (SC)  
... Det er ikke bare ulven som krysset bekken (OM)  
... Det er et større antall elger enn ganger ulven krysset bekken (CE)

4 Flere harer har hoppet over gjerdet enn geita har  
... Geita har færre harer enn de som hoppet over gjerdet. (CD)  
... Det er et større antall harer enn ganger geita hoppet over gjerdet (CE)  
... Haren hoppet over gjerdet flere ganger enn geita (SC)  
... Det er ikke bare geita som hoppet over gjerdet (OM)

5 Flere elger har løpt gjennom skogen enn jærv har  
... Jærv har færre elger enn de som løp gjennom skogen. (CD)  
... Det er et større antall elger enn ganger jærv løp gjennom skogen (CE)  
... Det er ikke bare jærv som løp gjennom skogen (OM)  
... Jærv løp gjennom skogen oftere enn elgen. (SC)

6 Flere mus har vært i åkeren enn rotta har.  
... Rotta har færre mus enn de som var i åkeren (CD)  
... Det er ikke bare rotta som var i åkeren (OM)  
... Musa var oftere i åkeren enn rotta. (SC)  
... Det er et større antall mus enn ganger rotta var i åkeren. (CE)

7 Flere pinnsvin har sett trestubben enn ekornet har.  
... Ekornet har færre pinnsvin enn de som så trestubben (CD)  
... Pinnsvinene så trestubben oftere enn ekornet (SC)  
... Det er et større antall pinnsvin enn ganger ekornet så trestubben (CE)

... Det er flere enn bare ekornet som så trestubben (OM)

8 Flere bjørner har gått i hi for vinteren enn grevlingen har.

... Bjørnene gikk i hi enn grevlingen har. (SC)

... Grevlingen har færre bjørner enn de som gikk i hi (CD)

... Grevlingen gikk færre ganger i hi enn antall bjørner (CE)

... Det er flere enn bare grevlingen som har gått i hi (OM)

9 Flere kjøttmeis har sittet på den øverste greina enn spurven har

... kjøttmeisene satt mer på øverste grein enn spurvene (SC)

... Spurvene har færre kjøttmeis enn de som satt på øverste grein (CD)

... Det er flere enn spurven som satt på øverste grein (OM)

... Det er et større antall kjøttmeis enn ganger spurven satt på greina (CE)

10 Flere rådyr har vært på vidda enn reinsdyret har

... Rådyrene var oftere på vidda enn reinsdyrene (SC)

... Det er et større antall rådyr enn ganger reinsdyret var på vidda. (CE)

... reinsdyret har færre rådyr enn de som var på vidda (CD)

... Det er ikke bare reinsdyret som var på vidda (OM)

11 Flere ulver har tatt et bytte enn jærv har.

... Ulven tok flere bytter enn jærv (SC)

... Det er et større antall ulver enn ganger jærv tok et bytte (CE)

... Det er flere enn bare jærv som tok et bytte (OM)

... jærv har mindre ulver enn de som tok et bytte (CD)

12 Flere lemen har opplevd vinteren enn spissmusa har.

... Lemenet opplevde vinteren flere ganger enn spissmusa. (SC)

... Det er ikke bare lemenet som opplevde vinteren (OM)

... spissmusa har færre lemen enn de som opplevde vinteren (CD)

... Det er flere lemen enn ganger spissmusa opplevde vinteren. (CE)

13 Flere elger har vært ute i snøen enn hjorten har

... Elgene var mer ute i snøen enn hjorten. (SC)

... Det er flere enn bare hjorten som var ute i snøen (OM)

... Det er et større antall elger enn ganger hjorten var i snøen (CE)

... hjorten har færre elger enn de som var ute i snøen (CD)

- 14 Flere spissmus har kastet stein enn ekornet har  
 ... Det er et større antall spissmus enn ganger ekornet kastet stein(CE)  
 ... Ekornet har flere spissmus enn de som kastet stein (CD)  
 ... Spissmusene kastet flere steiner enn ekornet(SC)  
 ... Det er ikke bare ekornet som kastet stein (OM)
- 15 Flere pinnsvin har sett mennesker enn ekornet har  
 ... Det er et større antall pinnsvin enn ganger ekornet så mennesker (CE)  
 ... Ekornet har færre pinnsvin enn de som så mennesker (CD)  
 ... Det er flere enn bare ekornet som så mennesker (OM)  
 ... Pinnsvinene så mennesker oftere enn ekornet (SC)
- 16 Flere katter har sovet under åpen himmel enn hunden har  
 ... Det er et større antall katter enn ganger hunden sov under åpen himmel (CE)  
 ... Kattene sov under åpen himmel oftere enn hundene (SC)  
 ... Hunden har færre katter enn de som har sov under åpen himmel (CD)  
 ... Det er flere enn bare hunden som sov under åpen himmel (OM)
- 17 Flere rever har snust i lyngen enn hunden har  
 ... Det er et større antall rever enn ganger hunden snuste i lyngen (CE)  
 ... Reven snuste i lyngen oftere enn hunden (SC)  
 ... Det er flere enn bare hunden som snuste i lyngen (OM)  
 ... Hunden har færre rever enn de som snuste I lyngen (CD)
- 18 Flere sauer har sett utsikten fra toppen av fjellet enn geita har  
 ... Det er et større antall sauer enn ganger geita så utsikten (CE)  
 ... Det er flere enn bare geita som så utsikten (OM)  
 ... Geita har færre sauer enn de som så utsikten (CD)  
 ... Sauen så utsikten oftere enn geita (SC)
- 19 Flere spurver har lagt egg i reiret enn skjæra har  
 ... Det er flere enn bare skjæra som la egg (OM)  
 ... Skjæra har færre spurver enn de som la egg (CD)  
 ... Spurvene la flere egg mer enn skjæra (SC)  
 ... Det er et større antall spurver enn ganger skjæra la egg (CE)

20 Flere rever har besøkt bondegården enn grevlingen har  
... Det er flere enn bare grevlingen som besøkte bondegården (OM)  
... Grevlingen har færre rever enn de som besøkte bondegården (CD)  
... Det er et større antall rever enn ganger grevlingen besøkte bondegården (CE)  
... Reven besøkte bondegården mer enn ekornet (SC)

21 Flere sauer har hilset på hesten enn kalven har.  
... Det er flere enn bare kalven som hilste på hesten (OM)  
... Sauen hilste på hesten ofter enn kalven (SC)  
... Kalven har færre sauer enn de som hilste på hesten (CD)  
... Det er et større antall sauer enn ganger kalven har hilset på hesten (CE)

22 Flere griser har rullet seg i søla enn villsvinet har.  
... Det er flere enn bare villsvine som rullet seg i søla (OM)  
... Grisene rullet seg I søla mer enn villsvinet (SC)  
... Det er et større antall griser en ganger villsvinet rullet seg i søla (CE)  
... Villsvinet har færre griser enn de som rullet seg I søla (CD)

23 Flere padder har besøkt dammen enn frosken har.  
... Det er flere enn bare frosken som besøkte dammen (OM)  
... Det er et større antall padder enn ganger frosken besøkte dammen (CE)  
... Frosken har færre padder enn de som besøkte dammen (CD)  
... Paddene besøkte dammen mer enn frosken(SC)

24 Flere katter har stukket hjemmefra enn hunden har.  
... Det er ikke bare hunden som stakk hjemmefra (OM)  
... det er flere katter enn ganger hunden stakk hjemmefra (CE)  
... Kattene har stukket hjemmefra oftere enn hundene. (SC)  
... Hunden har færre katter enn de som stakk hjemmefra. (CD)

25 Flere kaniner har funnet gulrøtter i åkeren enn haren har  
... Haren har færre kaniner enn de som har funnet gulrøtter. (CD)  
... Kaninene har funnet mer gulrøtter enn harene (SC)  
... Det er flere kaniner enn gulrøtter som haren fant (CE)  
... Det er ikke bare haren som har funnet gulrøtter (OM)

26 Flere muldvarper har gravet huler enn lemenet har  
... Det er ikke bare lemenet som gravde huler (OM)  
... Det er et større antall muldvarper enn ganger lemenet gravde huler. (CE)  
... Mudvarpene gravde flere huler enn lemenet (SC)  
... Lemenet har færre muldvarper enn de som gradve huler (CD)

27 Flere skjærer har stjålet sølv enn kråka har  
... Skjærene stjal mer sølv enn kråka (SC)  
... Kråka har færre skjærer enn de som stjal sølv (CD)  
... Det er ikke bare kråka som stjal sølv (OM)  
... Det er et større antall skjærer enn ganger kråka stjal sølv (CE)

28 Flere kaniner har knasket på gulrota enn haren har  
... Det er et større antall kaniner enn ganger haren knasket på gulrota (CE)  
... Det er ikke bare haren som knasket på gulrota (OM)  
... Haren har flere kaniner enn de som knasket på gulrota (CD)  
... Kaninene knasket på gulrota oftere enn haren (SC)

Filler sentences:

A. Reven trodde han ville ta musa raskere enn hva han gjorde  
... Reven fikk fanget musa X  
... Reven fikk ikke fanget musa  
... Reven var treigere enn musa  
... Rotta var raskere enn alle sammen.

B. Flere ekorn løp over gårdstunet sist onsdag enn røyskatter  
... Det var flere ekorn enn røyskatter X  
... Det var flere gårdstun enn ekorn  
... Ekornene hadde flere røyskatter  
... Røyskattene hadde flere ekorn enn gårdstun

C. Flere ender badet i dammen enn duer i parken  
... Det var flere ender enn duer

- ... Det var flere duer enn ender
- ... Det var flere dammer enn det var duer
- .. Det var flere duer enn det var parker

- D. Det var ingen andre enn en hvit hare som hadde lagd spor i snøen
- ... Den grå haren hadde laget spor i snøen
- ... Ingen hadde lagt spor i snøen.
- ... En hvit hare hadde lagd spor i snøen.
- ... Flere dyr hadde lagt spor i snøen.

- E. Flere ulver var i skogen enn det var hjort på fjellet
- ... Det var ingen hjorter på fjellet
- ... Det var noen hjort i skogen
- ... Det var flere ulver enn hjort
- ... Det var flere hjort enn ulver

- F. Noen harer løp lengre enn hjorten
- ... Det var kun hjorten som løp så langt
- ... Alle harene løp en kortere distanse enn hjorten
- ... Hjorten kunne ikke løpe så fort som harene kunne
- ... Noen harer løp en kortere distanse enn hjorten

- G. Reven så ulven jage sauene over haugene
- ... Reven jagde sauene lengre enn ulven
- ... Ulven jagde sauene lengre enn reven
- ... Det var kun ulven som jagde sauene
- ... Det var kun reven som jagde sauene

- H. Musa ville heller leke med muldvarpen enn å sove gjennom vinteren
- ... Musa klarte ikke å sove seg gjennom hele vinteren
- ... Musa fikk enkelt til å sove seg gjennom hele vinteren
- ... Muldvarpen sov seg gjennom vinteren lettere enn hva Musa gjorde.
- ... Musa ville ikke sove gjennom vinteren.

I. Bjørnen prøvde å få tak i honningen i bikuben, men endte opp med å bli jaget vekk av bier.

- ... Bjørnen tok mer honning enn hva biene tillot han å gjøre
- ... Bjørnen fikk ikke lov til å ta honning fra bikuben i det hele tatt
- ... Bjørnen fikk lov å biene til å ta honning fra bikuben
- ... Bjørnen fikk tak i så mye honning som han trengte.

J. Snøuglen hadde større øyne enn alle de andre fuglene i skogen

- ... De andre fuglene kunne ikke se snøuglen
- ... De andre fuglene hadde mindre øyne enn snøuglen
- ... Snøuglen hadde mindre øyne enn noen fugler
- ... Snøuglen kunne se bedre enn de fleste fuglene i skogen.

K. Musa synest at ost smakte mye bedre enn nøttene ekornet hadde samlet

- ... Ekornet hadde samlet sammmen masse ost for musa.
- ... Musa foretrakk ost fremfor nøttene til ekornet
- ... Ekornet hadde samlet nøtter kun for musa sin del.
- ... Musa foretrakk nøtter fremfor osten til ekornet.

L. Oksen løp raskere enn den stakkars mannen som gikk inn i garden.

- ... Mannen klarte å rømme fra oksen.
- ... Mannen klarte ikke å rømme fra oksen.
- ... Mannen løp raskere enn oksen.
- ... Mannen løp ikke raskere enn oksen.

M. Beveren hadde laget en demning som var større enn hullet muldvarpen hadde gravet

- ... Muldvarpen hadde laget et veldig stort hull
- ... Beveren hadde laget en veldig stor demning.
- ... Demningen og hullet var like store.
- ... Demningen var større enn hullet.



N. Oteren hadde gnaget mer på stokker enn på næringsrik mat den siste tiden

- ... Oteren hadde ikke fått spist på en lang stund
- ... Oteren hadde kun spist treverk på en lang stund
- ... Oteren hadde både spist mat og gnagd på stokker
- ... Oteren hadde kun spist næringsrik mat.

O. Biene hadde samlet mer honning i år enn hva de hadde gjort i fjor

- ... Biene samlet like mye honning begge årene.
- ... Biene samlet mer honning i fjor.
- ... Biene samlet mer honning i år.
- ... Biene samlet ikke noe honning noen av årene.

P. Anden hadde svømt over dammen flere ganger enn hva frosken hadde

- ... Anden og frosken hadde svømt over dammen like mange ganger
- ... Anden har svømt flere ganger enn frosken
- ... Frosken har svømt flere ganger enn anden
- ... Ingen av dem har svømt over dammen.

Q. Beveren hadde fanget mer fisk enn hva oteren hadde

- ... Oteren hadde ikke funnet noen fisker
- ... Beveren hadde ikke funnet noen fisker
- ... Oteren hadde ikke fanget like mange fisker som beveren
- ... Beveren hadde ikke fanget like mange fisker som oteren.

R. Padden hadde fått flere barn i fjor enn hva frosken hadde

- ... Padden og frosken hadde like mange barn.
- ... Frosken hadde færre barn enn hva padden hadde.
- ... Padden hadde færre barn enn hva frosken hadde.
- ... Hverken padden eller frosken hadde fått noen barn.

S. Hjorten hadde et større givir enn alle reinsdyrene han noensinne hadde

sett

- ... Noen reinsdyr hadde større givir enn hjorten
- ... Hjorten hadde det største giviret.
- ... Reinsdyret hadde det største giviret.
- ... Det fantes hjort som ikke hadde et givir.

T. Ulven var mer redd for mannen enn hva mannen var redd ulven.

- ... Mannen var mer redd ulven
- ... Ulven var mer redd mannen
- ... Mannen og ulven var like redde hverandre.
- ... De var ikke redde hverandre.