



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„The meaning and cognition of irony“

verfasst von / submitted by

Susanne Veil BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien, 2016 / Vienna 2016

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 066 867

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Allgemeine Linguistik: Grammatiktheorie und
kognitive Sprachwissenschaft

Betreut von / Supervisor:

Ao.Univ.-Prof. Dr. Chris Schaner-Wolles

Abstract

This thesis aims to describe irony within linguistics from a cognitive perspective. Classic and recent theories and frameworks on figurative language and humor are described and compared. They are then discussed in the light of current psycho- and neurolinguistic studies that provide evidence for irony as a two-step process that depends on various clues and is deeply linked to theory of mind. For irony to be successful it finally comes down to the question whether the hearer is able or willing to assume that the speaker is cooperative, this means that the hearer assumes that the speaker knows that the hearer understands. These are called second-order beliefs that have proven to be essential for understanding irony. Finally, new possible insights for irony within experimental pragmatics are discussed and questions posed for further research.

Acknowledgements

I like to thank my supervisor Ao.Univ.-Prof. Dr. Chris Schaner-Wolles for her patience to check my thesis over and over again, not overlooking even the smallest details. Furthermore, I thank Ass.-Prof Mag. Dr. Hans Martin Prinzhorn for his advice and insight. I also owe a huge part of the argument about how the enrichment of meaning of irony could be modeled to Daniele Panizza, PhD.

Contents

1	Introduction	1
2	Pragmatic theories	4
2.1	Grice	6
2.1.1	The standard pragmatic model	11
2.1.2	Criticism	12
2.2	Relevance	12
2.2.1	Joint pretense	16
2.2.2	The direct access view	17
2.2.3	Criticism	18
2.3	Neo-Gricean accounts	19
2.3.1	The graded salience hypothesis	19
2.3.2	Relevant inappropriateness	22
2.3.3	Criticism	25
2.4	General theory of verbal humor	26
2.5	Presuppositions	31
2.5.1	Common ground	36
2.6	Discussion	42
2.6.1	Implicatures and irony	42
2.6.2	Presupposition and irony	48
2.6.3	Irony as enrichment of meaning	52
3	Psycho- and neurolinguistic studies	61
3.1	Two-step interpretation	62
3.2	Theory of mind	65
3.3	Communicative intentions	75
3.3.1	Communicative intentions and Autism	80
3.4	Discussion	86
4	Situational and computational irony	93

5	General discussion	98
6	Conclusion	104

Chapter 1

Introduction

We live in a time where this figure of speech is so widespread that some even used it to title an entire epoch: It was said that irony was the figure of speech of postmodern times (Gibbs, 2000, pg 5). In 2000 the novelist Dave Eggers wrote about irony in the appendix of his memoir in an own section "Irony and its Malcontents": "It is without doubt the most over-used and under-understood word we currently have" (Eggers, 2000, pg 33). Today some even claim that there exists such a thing like "'post-modern' modes of irony" (Wallace, 2015, pg 480), where the ironist does not assert any true meaning at all, but merely distances him- or herself from whatever issue.

Now that a post-modern use of irony seems to have evolved, have we understood "classical" instances of irony? This thesis aims to contribute to the vast literature on irony a more narrow, linguistic perspective. Since it is believed here that we can only understand a linguistic concept that widely used if we consider it a formal question that needs to be answered in technical terms.

I intend to do so by asking the question: what does ironic speech mean? Rather than asking: how is it used? Central is the idea that we have to understand its meaning before we can grasp how it is used. We have to get from meaning to usage, and we cannot infer the meaning from the usage. This is actually what Searle (1969) already pointed out to in his speech act theory speaking about the fallacy of meaning as use:

The truth conditions of the one proposition may be sufficient for the truth conditions of the other – even though the point of uttering one sentence may be different from the point of uttering the other sentence. The truth conditions of a proposition have been

confused with the point of force of uttering a sentence, because the word 'use' is so vague as to include both the truth conditions of the propositions expressed and the point or illocutionary force of uttering the corresponding sentence (Searle, 1969, pg 148).

Irony is only possible with a certain context, which makes it a pragmatic issue. We will therefore try to understand how irony interacts with the context in which it is uttered.

In the first chapter, I will outline previous approaches to irony and its comprehension that are concerned on making statements about its meaning. I will discuss them at length touching issues that are always crucial for pragmatics, like implicatures and presupposition. This leads to a discussion about those theories on how the meaning of irony could be described. Then, I present psycho- and neurolinguistic studies focusing on the use of irony and humor conducted among the following populations: children, brain damaged patients and participants from the Autistic Spectrum Disorder. I will sum up contradictory and consistent findings and discuss the meaning of those for the theories. Finally, I will conclude with some remarks on situational irony that shows how humor and irony interact and why we perceive contradiction as funny. In the same chapter, I will briefly outline some new approaches on how to model irony computationally. I will complete with a discussion that brings the theories, empirical findings, and similarities between irony and jokes together.

Definitions At the beginning, some definitions are in order. Irony, as mentioned in the beginning, is a very widely used concept and correspondingly wide are the definitions employed.

I will use the definition by Zimmermann and Sternefeld (2013): "the conveyed meaning is exactly the **opposite** of the literal meaning" (Zimmermann and Sternefeld, 2013, pg 5, boldface in the original). They also mention that:

In classical rhetoric, irony is always defined as expressing the opposite of the literal meaning. In ordinary language, however, the term *irony* is used in a much broader sense (Zimmermann and Sternefeld, 2013, pg 5, italics in the original).

There is no doubt that in colloquial language the term *irony* can be used to convey the whole range of possible meanings from the opposite of what has been said, to only slight variations between what is uttered literally and what is meant. This is also expressed in psycholinguistics, here I like to quote

as an example the study by Nakassis and Snedeker (2002, pg 429), they state: "Irony is a gradient phenomenon, ranging from subtle to crass [...]". That is without doubt true for everyday language use. Nevertheless, I will restrict myself to the narrow definition as it is common for scientific reasons to focus on the most distinct and therefore stereotypic form of a phenomenon to use it as exemplary and describe its features that can be used then to apply to instances of the phenomenon farer away from its stereotypical form.

It is also necessary to put irony, as defined now, in relation to different shades of non-literal, jocular language such as sarcasm and cynicism. I would argue that one can distinguish them like this:

- Irony: The opposite of the literal meaning.
- Sarcasm: In contrast to irony, sarcasm cannot be pinpointed by its meaning alone but is defined by its pragmatic usage. Sarcastic utterances convey a critical attitude towards whatever is expressed. Therefore I would say that sarcasm is a subcategory of irony that always expresses critical feelings towards the issue talked about.
- Cynicism: Cynicism conveys a general criticism of social norms and moral rules. Therefore this instance of non-literal speech can be used to describe a person: Calling someone a *cynic* expresses something about a persons general attitudes towards social norms and therefore is understood as saying something about the more general characteristics of a person.

That these definitions are hypotheses at the same time will become clear in the next chapter, since the Gricean idea of irony conveying the opposite of the literal meaning has been questioned among more recent theories on non-literal meaning.

All examples that are not marked otherwise, are my own, picked up in communication.

Chapter 2

Pragmatic theories

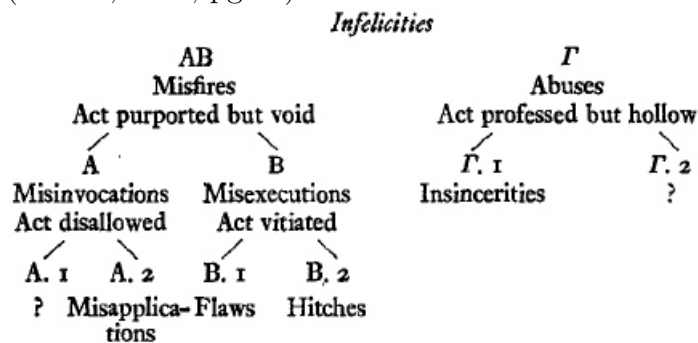
The theories presented here are selected because they have proven to be very influential in humor research. And also because they employ roughly the notion of irony communicating the opposite of what is said (see Brône, 2012, pg 485).

In general, there have been two lines of theorizing about irony: First, the theories that understand irony as Gricean trope. That is to say that **irony is a figure of speech** to express something that cannot be expressed this efficiently via a literal interpretation. This is done by flouting the maxim of quality (Grice, 1967, pg 49). This line of thinking has been widened in so far as that for Neo-Gricean approaches it can be every other maxim as well (see Attardo, 2000). Second, the theories that understand **irony as mention or pretense** that conveys sentiments toward an echoed issue. Therefore irony expresses feelings or sentiments relating to an issue or a thing or a person touched upon in the previous dialogue. But the victim has to be present and salient to be the anchor of the ironic utterance (see Sperber and Wilson, 1981). The anchor is the background to use for the comparison of the ironic with the literal interpretation to achieve the most relevant interpretation.

I will review and discuss the Gricean, Neo-Gricean and relevance-theoretic account, as well as a few other pragmatic approaches on irony and humor that I consider very interesting. A question I try to answer in discussing these approaches is, which assumptions the different theories make in terms of the distinction between irony and implicature. Before doing that, I would like to review very shortly the theories by Austin (1962) and Searle (1969) that have had huge influence on pragmatics, and are therefore the background for the theories reviewed here. Also, in their theories there is the idea for something like irony present, where one means something different than what is said.

Austin's felicity conditions In contrast to "constatives" (Austin, 1962, pg 3) that are made to report reality, Austin (1962) introduced his famous "performative sentence[s]", when it "is not to *describe* my doing of what I should be said in so uttering to be doing or to state that I am doing it: it is doing it" (Austin, 1962, pg 6, italics in the original). To perform a performative utterance, according to Austins **felicity conditions**, the speaker has to follow a conventionalized procedure, and has to be qualified to do that, and he or she has to do that right and completely (Austin, 1962, pg 14, 15). Furthermore he or she has to mean it and behave accordingly to what he or she said. Are his or her feelings not in line with the performative act, that is the last two conditions are not obeyed, Austin calls that break of rules an "abuse" (Austin, 1962, pg 16). By abusing the performative act, one is a liar, or at least insincere (Austin, 1962, pg 18). The act is not void, but Austin states, talking about the speakers thoughts, that there is a parallel element to lying to that, "in performing a speech-act of an *assertive* kind" (Austin, 1962, pg 40, italics in the original). Austin speaks about the etiulative use of language, when there is a mutual abuse obvious for the interlocutors: "Language in such circumstances is in special ways – intelligibly – used not seriously, but in ways parasitic upon its normal use – ways which fall under the doctrine of the *etiulations* of language" (Austin, 1962, pg 22, italics in the original).

(1) (Austin, 1962, pg 18)



Austin calls the cases where something goes wrong with the "smooth or 'happy' functioning of a performative" (Austin, 1962, pg 14) **infelicities**, that are "the doctrine of *the things that can be and go wrong*" (Austin, 1962, pg 14, italics in the original). While not obeying the first four felicity conditions among the six, makes an utterance void and annuls it, the infelicity of the last two rather weakens an utterance. Austin calls that an "abuse of the procedure" (Austin, 1962, pg 25). Where acts are "professed but hollow", one possibility is that they are "Insincerities" and the other possibility is "?".

Maybe in (1) Γ is where jocular language can be located in Austins framework of performatives. That might be inexplicit performatives, since I do make a joke instead of describing one, when I tell a joke or I am being ironic instead of describing irony.

Searle's primary illocutionary acts In his Theory of Speech Acts Searle (1969) asked in the beginning: "What is the difference between saying something and meaning it and saying it without meaning it?" (Searle, 1969, pg 3). He is concerned with this question when later writing about **indirect speech acts**:

In indirect speech acts the speaker communicates to the hearer the more than he actually says by way of relying on their mutually shared background information, both linguistic and nonlinguistic, together with the general powers of rationality and inference in the part of the hearer (Searle, 1975, pg 60,61).

In Searle's view it is the **primary illocutionary act** that is only implicit and has to be computed inferentially. In order to do that, the hearer needs an inferential strategy that tells him when second and literal illocutionary act are not the same. This strategy requires mutual background information, a theory of speech acts, and general principles of cooperative conversation (Searle, 1975, pg 63,64). The cooperative principle gets expanded or differentiated by two features: First, "a strategy for establishing the existence of an ulterior illocutionary point beyond the illocutionary point contained in the meaning of the sentence" (Searle, 1975, pg 74), established by the principles of conversation, and second, a device for finding out what this point is, derived from the theory of speech acts plus background information.

No one has had bigger influence in describing the principles that govern conversation than H. P. Grice, which is why his theory and his ideas on irony are the first to be discussed here.

2.1 Grice

In his famous text "Logic and Conversation" Grice (1967) explored the fact that in conversation we can hint on, suggest, or imply something different or more than what is actually said. Grice speaks of the intuitive meaning of *said* here and means the strict lexical meaning of the words themselves, "what someone has said" which is "closely related to the conventional meaning of the words [...] uttered" (Grice, 1967, pg 44). What is implied and not said,

Grice calls "implicature" (Grice, 1967, pg 44). Some of these implicatures are such that by the use of certain words they automatically bring an additional meaning with them, those are "conventional implicatures", associated with the "conventional meaning" (Grice, 1967, pg 44) of what is literally said. This stands in contrast to those "conversational implicatures" (Grice, 1967, pg 45) that depend highly on the language context and the situation in which a sentence is uttered. This is the case in the following example (Grice, 1967, pg 43):

- (2) A asks B how C is getting on in his job, and B replies, *Oh quite well, I think; he likes his colleagues, and he hasn't been to prison yet.*

For some the sentence above can also be understood as an jocular ironic utterance, which shows already the central question we have to answer here: Where do (conversational) implicatures end and where does irony begin? Or can one say: since B does not want to say that C has actually been sent to prison he does not convey the opposite and is therefore not being ironic? I will come back to this question at the end of this section in (2.6.1).

According to Grice (1967, pg 45) **conversational implicatures** are a subclass of nonconventional implicatures that are connected with the following features of the current discourse: The conversational maxim of quantity (be as informative as necessary), quality (be true and do not say something that you lack evidence for), relation (say what is relevant), and manner (be concise). In sum, they serve the **cooperative principle** (CP): "Make your conversational contribution such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" (Grice, 1967, pg 45). For Grice, talking is a rational behavior that succeeds because we obey to a shared set of rules, even though he acknowledges that the most effective communicational exchange is an ideal that we rarely achieve (Grice, 1967, pg 47). And in some cases we willingly fail to obey all the maxims. Among the ways to fail a maxim there is the possibility to "flout" (Grice, 1967, pg 49) one, in other words, "BLATANTLY fail to fulfill it" (Grice, 1967, 49, stress in the original). If the reason for that isn't that the speaker is unable to obey the maxim, neither is he or she avoiding the clash between two competing maxims, nor is he or she trying to lie, than the hearer is assuming that the speaker nevertheless obeys the cooperative principle. In this case, we are talking about a conversational implicature that is "exploiting" (Grice, 1967, pg 49) a maxim. To decode a conversational implicature the hearer needs the following information: (1) the literal meaning of the words themselves and their references; (2) the cooperative principle and the maxims; (3) the linguistic and non-linguistic context of what has been said; (4) background knowledge; (5) the knowledge that this all is not

only known by the hearer but also by the speaker, and that both know or at least assume that mutual knowledge is the case. (Grice, 1967, pg 50) So we receive the following structure for conversational implicatures:

He has said that p ; there is no reason to suppose that he is not observing the maxims, or at least the CP; he could not be doing this unless he thought that q ; he knows (and knows that I know that he knows) that I can see that the supposition that he thinks that q is required; he has done nothing to stop me thinking that q ; he intends me to think, or is at least willing to allow me to think, that q ; and so he has implicated that q (Grice, 1967, pg 50).

Since the conversational maxims and the conversational implicatures are connected, disobedience of one of the conversational maxims can give rise to a conversational implicature (Grice, 1967, pg 47). One of Grice's examples that involve the exploitation of the maxim of quality is irony. His example (Grice, 1967, pg 53):

- (3) X, with whom A has been on close terms until now, has betrayed a secret of A's to a business rival. A and his audience both know this. A says '*X is a fine friend*'.

This example differs from the one in (2) in that here the speaker A obviously does not obey the maxim of quality and says something that is wrong. The cooperative principle still stays intact if the hearer assumes that the speaker knows that A does not want to communicate the literal meaning. Rather the speaker wants to communicate some other proposition. Grice writes: "This must be some obviously related proposition; the most obviously related proposition is the contradictory of the one he purports to be putting forward" (Grice, 1967, pg 53).

A few years later Grice added the following to the issue of irony being conversational and non-conventional implicatures: "[...] irony is intimately connected with the expression of a feeling, attitude, or evaluation. I cannot say something ironically unless what I say is intended to reflect a hostile or derogatory judgment or a feeling such as indignation or contempt" (Grice, 1978, pg 124). We will come back to this idea in section (2.2) about pretense and the echoic mention theory.

To distinguish conversational from conventional implicatures, Grice (1967) adds that the former is said in a particular situation from which we derive a certain context. Furthermore, to compute the meaning of a conversational implicature one has to calculate what is needed plus the utterance to obey

the cooperative principle; that is, the supposition plus the literal meaning. This chain of processing is close to, as we will see in just a bit, the retention hypothesis. As the possible explanations that the hearer employs to make sense of an utterance may be various, the hearer can not know with certainty which one is the case, there remains always a certain amount of uncertainty, or as Grice calls it "indeterminacy" (Grice, 1967, pg 58). Another feature that Grice (1967, pg 57) points out to for generalized conventional implicatures is that they can be canceled. This can be done explicitly by saying something that deletes the implicature, or implicitly via a context that makes clear that the implicature is rejected in this case.

Grice puts conventional implicatures right away outside of the explanatory realm of the cooperative principle, since they are derived by the meaning of the words themselves. As I stated above the conventional meaning influences not only what is literally said but also what is implied conventionally. The following is indeed a strange passage, as Potts (2005, pg 9) points out to:

In some cases the conventional meaning of the words used will determine what is implicated, besides helping to determine what is said (Grice, 1967, pg 44).

Potts (2005) argues that this means Grice understands conventional implicatures as being inside the grammar and not part of pragmatics. While conversational implicatures or presuppositions can and should be covered by pragmatic principles, conventional implicatures cannot (Potts, 2005, pg 9). That conversational implicatures are inherently pragmatic is also confirmed by one of the features Grice gives them:

Since the truth of a conversational implicatum is not required to be the truth of what is said (what is said may be true – what is implicated may be false), the implicature is not carried by what is said, but only by the saying of what is said, or by 'putting it that way' (Grice, 1967, pg 58).

To apply this to irony: What is implicated ironically can be true even when what is said may be false, just like in (3), where it is true that X is a bad friend. Grice (1967, pg 58) speaks of "nondetachability" for implicatures that is high for generalized conversational implicatures, meaning they have to be "carried by a familiar, nonspecial locution" which is certainly not the case for irony. Using this terminology would mean irony has a high degree of detachability (or low degree of nondetachability), since there is apparently no locution that is favored among others for ironic use.

As Potts (2005, pg 26) argues it is easy to distinguish conventional from conversational implicatures since Grice used the terminology in a way to construct an opposite relationship between the two. Conversational implicatures as noted above use the maxims and the cooperative principle in a way where, if taken for granted, their disobedience communicates something. Conventional implicatures, on the other hand, use the common meaning of the words themselves. Grice's example for an conventional implicature reads as follows (Grice, 1967, pg 44).

- (4) I say (smugly): *He is an Englishman; he is, therefore, brave*

Grice argues that by saying it I imply that the word *Englishman* implies bravery. By arguing as strangely as he does here, in a way using double-implication, it is no wonder that others (for example Reimer (2013)) have argued that Grice says understanding language involves a second-order-interpretation. But we will come to that when talking about studies on theory of mind in (3.2) that might support this claim.

What is suspicious here, is the "smugly". If I can say that with this kind of emotion, can't I use a conventional implicature like this to convey criticism in an ironic way: Saying someone is an Englishman while meaning someone has bad manners, flouting the conventional meaning of the word? There is only one short remark of Potts (2005, pg 18) on the relation between conventional implicatures and irony:

- (5) a. Edna is at her friend Chuck's house. Chuck tells her that he thinks all his red vases are ugly. He approves of only the blue ones. He tells Edna that she can take one of his red vases. Edna thinks the red vases are lovely, selects one, and returns home to tell her housemate, 'Chuck said I could have one of his lovely vases!' (Potts, 2005, pg 18, stress in the original)
- b. 'Chuck said I could have one of his lovely vases. *But they are all so ugly!' (Potts, 2005, pg 28)

So Chuck said that Edna could have one of his red vases, and the *lovely* as an conventional implicature implies in a. that Edna thinks they are lovely. A continuation of this sentence as in b. is only possible when this conventional implicature-reading disappears. For example, because the *lovely* is part of what Chuck said to Edna about those vases and she is merely repeating it. Potts (2005, pg 28) suggests that in this case an ironic or sarcastic meaning is likely. When the opposite is meant, the continuation makes sense. This means that when the conventional implicature disappears irony becomes a possible meaning. Uttering *He is an Englishman and he is therefore brave*,

thus, can either imply that there is a connection between being an Englishman and bravery (being an Englishman entails being brave), or it is ironic, but not both at the same time. Thus, if meant ironically then the conventional implicature as form of entailment disappears. This makes sense, since conventional implicatures and irony have to be detachable, while conversational implicatures have the feature of "nondetachability" (Potts, 2005, pg 27) to a certain construction. Apparently, either there is a meaning, like that of an conventional implicature, attached to a certain form of an utterance, or this meaning gets lost in favor of a more generalized meaning like irony.

2.1.1 The standard pragmatic model

The so-called **standard pragmatic view** is based on Grice's definition of irony as conversational implicature and states that irony is a figure of speech that communicates the opposite of what is said. Also, Grice mentioned that in order to fully conduct the ironic meaning, we have to reconstruct what is missing for an utterance to obey the cooperative principle. Therefore only if we as hearer are aware of the literal unintended meaning we can appreciate the irony as being different from it. This is described in the standard pragmatic view, where it is assumed that the comprehension of non-literal language takes place in a two-stage process: First the listener must compute the utterance's context-independent, literal interpretation, if a mismatch with the context indicates that the literal interpretation is inappropriate it is necessary to cancel the surface-literal interpretation and compute the non-literal interpretation. Overall, this means further processing effort and higher processing cost for non-literal than for literal language. This point of view relates to Grice, who used irony as an example for conversational implicatures where the hearer needs to know what is missing of an utterance for it to still obey the cooperative principle. His account is pragmatic in that irony is considered a form of violation of one of the maxim of quality and evades the cooperative principle (Brône, 2012, pg 485). This two-stage process will be discussed at more length in the section on psycholinguistic studies in (3.1). A more recent approach that also assumes that comprehension of irony is about knowing what is needed to repair the literal meaning of an utterance in order to make it obedient of the cooperative principle is the graded salience hypothesis that we turn to when discussing Neo-Gricean accounts in (2.3).

2.1.2 Criticism

Coming back to the example in (2), the understatement in it can also be understood as ironic, because it implies that speaker B obeys the cooperative principle and gives the most information he can about how C is doing at his new job. If the most one can say about his success is that he has not been arrested yet, this is not much. The *yet* implies that he could still go to prison because his colleagues aren't nice at all but rather mean and "treacherous" (Grice, 1967, pg 43) people, this is the possible interpretation actually provided by Grice. In any case, the speaker means something different than he says. The example in (2) and the one in (3) are both conversational implicatures in Grice's definition.

So both examples (2) and (3) fulfill the conversational principle, but in the first one the speaker very neatly obeys the maxim of quantity, in saying exactly as much as he knows to be the truth, while in the second one the speaker very obviously does not obey the maxim of quality, and says something that all interlocutors know to be false, which leads the hearers to assume it's an ironic comment.

Let's assume that implicatures can be ironic as in (3) or as in (2). For Grice irony is a special case of conversational implicatures, derived by flouting the maxim of quality. However, as I tried to show above, one can argue you can't only derive irony by the use of the maxim of quantity. You can also be ironic not by violating the maxim, but also by obeying it. Furthermore, I can also flout the conventional meaning of a word to derive irony, as I argued with the example of an Englishman without manners. Grice's definition of irony might be too narrow. There has been an extension of the Gricean account in that researchers assume that the violation of any maxim can trigger irony (Brône, 2012, pg 485). Widening the conditions of the use of conversational maxims in order to derive irony was the endeavor of theories that can be called Neo-Gricean.

2.2 Relevance

While Grice is the main proponent of a two-stage process for irony comprehension, the account of Sperber and Wilson (1981) and (tion) that I will present now is the main proponent for a direct-access or one-stage view for irony comprehension (Attardo, 2000, pg 797). Their theory is the most influential amongst theories of irony as mention or pretense (Brône, 2012, pg 489). It, therefore, stands in contrast to the theory of Grice and theories that built upon the ideas of Grice, not only because they make different predic-

tions in terms of the comprehension process of irony, but also because they understand irony as mention and not as figure of speech.

The point of departure for Sperber and Wilson (1981) is that they assume the traditional semantic account, that irony figuratively means the opposite of what literally was said, and the purely pragmatic account by Grice, both fail due to the same reasons: There is no reason why one would make the effort to communicate indirectly the opposite and not say it directly. Furthermore, both do not provide a mechanism that shows exactly how the hearer comes from the explicit literal meaning to the implicit ironic meaning. And as last point of criticism, there is no precise definition, neither of the semantic oppositeness, nor of the Gricean type of irony as an implicature and no explanation why irony should work like other standard cases of implicatures (Sperber and Wilson, 1981, pg 295-296).

I will first review the basic claims of their proposal and then go into the arguments they give for their criticism of a purely semantic account and against Grice's purely pragmatic account.

Their **echoic mention theory** is rooted in **relevance theory**. What they mean by "relevant" is described like this: "The propositional form of an utterance is an interpretation for a mental representation of the speaker which can be entertained as an interpretation of a desirable (e.g. relevant) representation" (Sperber and Wilson, 1981, pg 232). To make our "communicative intention" (Sperber and Wilson, 1981, pg 157) clear we chose the stimulus that is the most relevant, which means requires the least processing effort for the addressee because it does provide new information that can be integrated easily due to its relatedness to old information (Sperber and Wilson, 1981, pg 48). Irony is understood as a form of echoic utterance to express the reference to a previously encountered utterance and the speaker's attitude towards it, that can be of the "rejecting or disapproving kind" (Sperber and Wilson, 1981, pg 239).

To them irony is a "garden-path utterance, likely to cause the reader momentary processing difficulties later offset by appropriate rewards. [...] By leaving the echo implicit when the addition of some explicit material would have immediately put the reader on the right track, the author opens up a whole new line of interpretation" (Sperber and Wilson, 1981, pg 242).

According to this theory irony itself is an interpretation already: "It is an interpretation of a thought of someone other than the speaker (or of the speaker in the past) [... A] second-degree interpretation of someone else's thought" (Sperber and Wilson, 1981, pg 238). Therefore ironic utterances can be called echoic and as such they allow the speaker to express his or her interpretation towards the echoed thought of someone:

From the pragmatic point of view, what is important is that a speaker can use an echoic utterance to convey a whole range of attitudes and emotions, ranging from outright acceptance and endorsement to outright rejection and dissociation, and that the recognition of these attitudes and emotions may be crucial to the interpretation process (Sperber and Wilson, 1995, pg 240).

Sperber and Wilson (1981) assume that there are two types of irony: echoic irony and standard irony. But since there exists a whole range of nuance between the two they rather conclude that all instances of irony are echoic but the echoes come in different "degrees and types" (Sperber and Wilson, 1981, pg 309). I want to revise here two examples that Sperber and Wilson give for ironic use of echoic mention.

(6) (Sperber and Wilson, 1981, pg 307, 308)

- a. *'You take an eager interest in that gentleman's concerns', said Darcy in a less tranquil tone, and with a heightened colour. 'Who that knows what his misfortunes have been, can help feeling an interest in him?' 'His misfortunes! repeated Darcy contemptuously, 'yes, his misfortunes have been great indeed.'*
[Jane Austen; *Pride and Prejudice*]
- b. *Elinor looked at him with greater astonishment than ever. She began to think he must be in liquor; ... and with this impression she immediately rose, saying, 'Mr Willoughby, I advise you at present to return to Combe – I am at not at leisure to remain with you longer. – Whatever your business may be with me, it will be better recollected and explained tomorrow.'* *'I understand you,' he replied, with an expressive smile, and a voice perfectly calm, 'yes, I am very drunk. – A pint of porter with my cold beef at Marlborough was enough to over-set me.'*
[Jane Austen; *Sense and Sensibility*]

What happens here is described as follows: "The speaker mentions a proposition in such a way as to make clear that he rejects it as ludicrously false, inappropriate or irrelevant. For the hearer, understanding such an utterance involves both realizing that it is a case of mention rather than use, and also recognizing the speaker's attitude to the proposition mentioned. The whole interpretation depends on double recognition. Recovery of the implicatures [...] will follow automatically:" (Sperber and Wilson, 1981, pg 308)

(7) *He has not been the victim of misfortunes.*

(8) *I am not drunk.*

So Mr Dary and Mr Willoughby echo the previous utterances of Elizabeth and Elinor and in repeating them they also ridicule the echoed utterances. In doing so, both men reject the accusations, which means they are expressing their attitude towards the judgments brought forward against them.

The "strange asymmetry in the use of irony" Sperber and Wilson (1981, pg 312) observe: that we are much more likely to use a positive proposition, expressing ironically failure and criticism than the negative literal meaning to express praise, is explained by them in that norms and standards are always salient and positive, therefore, their echoing is more straightforward than the echoic mentioning of negative things (Sperber and Wilson, 1981, pg 312–315).

Sperber and Wilson (1981) propose their so-called **substitution theory of irony**, where they assume "that the interpretation of ironical utterances cannot be reduced to the search for conversational implicatures" (Sperber and Wilson, 1981, pg 299). Sperber and Wilson (1981) suggest an account that includes some semantic and pragmatic features to explain the usage of irony, where the implicature substitutes for a literal meaning instead of replacing it (Sperber and Wilson, 1981, pg 299). They propose that the speaker of an ironical utterance doesn't want to communicate the content of an utterance, but wants to communicate an attitude or a state of mind in a certain situation (Sperber and Wilson, 1981, pg 302–303). The difference between the sentences a. and b. and the sentences c. and d. in (9) is that the first two refer to an attitude to the weather while the last two refer to an attitude to the content of an utterance. The latter being some kind of metarepresentation. **Using** an expression "involves reference to what the expression refers to" (Sperber and Wilson, 1981, pg 303), while **mentioning** an expression "involves reference to the expression itself" (Sperber and Wilson, 1981, pg 303). This is meant if they say that irony is a form of mention not use. It refers to the metarepresentational use of irony, not referring to an entity in the world, like the weather, but to an utterance about the world, like *What lovely weather*.

(9) [Two people caught in a downpour]

a. *What awful weather.*

b. *It seems to be thundering.*

c. *What lovely weather.*

d. *It seems to be raining.* (Sperber and Wilson, 1981, pg 300)

According to the approach by Sperber and Wilson (1981) both sentences, c., and d. are considered ironic. Even though by saying d. the speaker

surely does not mean the opposite rather he or she is stating the obvious. So it would not be necessary to utter d. if I would want to inform my hearer about the fact that it rains. Rather I utter the sentence to make a personal comment about the weather. Therefore, irony allows me to express my feelings or attitudes in a sophisticated and funny way. Sophisticated because I do not express my emotion directly, it is echoed indirectly. Ironic utterances have a semantic property, they are mentions, to this an inferential pragmatic process must be supplemented in that those mentions are echoic. So semantics and pragmatics are combined when irony is understood as echoic mention (Sperber and Wilson, 1981, pg 316-315).

2.2.1 Joint pretense

I will only briefly mention that there exists also the **joint pretense view** by Clark and Gerrig (1984). According to this irony requires an additional pretense layer in discourse, rooted in the ironist's desire to express a critical attitude towards the issue in question. He or she doesn't want to do that openly but prefers a somewhat distanced, pretended form (Brône, 2012, pg 489, 490). Also, there exists the **allusional pretense view** as an adaptation of the reminder theory. Irony is seen as a form of reference to a currently violated but otherwise common norm. In the **mental spaces theory** irony is considered to be outsourced to a mental space, a created world, and the speaker makes this act of creation explicit, so the hearer doesn't assume he or she is lying in the real world (Brône, 2012, pg 490, 491).

Fairly obvious, these accounts are related to relevance theory. They pick out two different arguments of the relevance theoretic account and put them in the middle of the assumptions about irony. For the joint pretense view, this is the idea that irony allows the speaker to express his or her feelings towards the issue that he or she mentions ironically. This was the argument why Sperber and Wilson called irony a second-degree interpretation, here this is formulated as irony being a distanced form of criticism.

The same holds for the allusional pretense view, here the reason for irony is not to express a critical attitude, but the violation of common norms. This is also about irony being some kind of metarepresentation that Sperber and Wilson talked about. Since both ideas are so close to the echoic mention theory as part of relevance theoretic accounts, I do not discuss them separately rather I assume that in the discussion for them is the case what is the case for the subordinate theory.

2.2.2 The direct access view

There exists also the opposing view to the standard pragmatic model: The **direct access view** assumes that the ironic meaning can be accessed directly without accessing the literal meaning first. Therefore it is assumed that contextual information interacts with lexical processes very early on. Similar underlying mechanisms are involved in the initial processing of both literal and figurative language. Hence, it does not distinguish in processing of literal or figurative language. Understanding irony, therefore, does not necessarily require special cognitive processes beyond those used to comprehend literal speech. If a particular context supports an ironic interpretation of a statement, the required interpretation can be directly accessed without the need to access the literal interpretation first. So no extra processing steps are required and no extra processing cost is assumed (Filik and Moxey, 2010, pg 422). This is the processing assumed by relevance theoretic accounts since their general idea is that processing is primarily guided by what is relevant in a certain context, and no matter if this is the literal or non-literal interpretation, the relevant meaning should be accessed immediately.

In psycholinguistic research there has been evidence for both, the two-stage process and the direct access view (Brône, 2012, pg 485).

As an instance of psycholinguistic research I would like to mention the study by Filik and Moxey (2010). They conducted an eye-tracing study on the on-line processing of written ironic phrases. They investigated in the activation of patterns of pronominal reference. "[...] positive and negative quantifiers can lead to focus on different sets of discourse entities [...] positive quantifiers lead to focus on the reference set, whereas negative quantifiers make the complement set more available" (Filik and Moxey, 2010, pg 423). Since irony can communicate the opposite of what has been said, the activation patterns should be opposite for cases of understood irony compared to literal speech. Their results are more in line with the predictions of the standard pragmatic model and the graded salience hypothesis, that states that it depends on salience, which interpretation is computed first: For unfamiliar irony the literal meaning is computed first, while for familiar irony it is a direct access of the non-literal ironic interpretation. (Filik and Moxey, 2010, pg 430) The graded salience hypothesis comes together with the retention hypothesis that assumes that in a two-stage process not one interpretation is computed and then rejected in favor of the other interpretation, but both interpretations remain active and accessible (Filik and Moxey, 2010, pg 431). "[...] for ironic sentences, both interpretations are part of the reader's mental representation of the meaning of the text as they try to understand it" (Filik and Moxey, 2010, pg 430).

2.2.3 Criticism

About the distinction between irony and sarcasm Sperber and Wilson (1981) say, that it depends on the utterance being an echo of the speaker, for irony, or the hearer, for sarcasm. If the speaker echoes himself it is irony, if the speaker echoes the hearer it is sarcasm. And indeed self-irony most often is self-critique. That is why we find it so appealing, it is a way of showing modesty. But irony towards someone else isn't always criticism, I can also praise someone with irony (even though this is more seldom than criticism):

(10) Indeed, you are a terrible singer.

If someone just surprised me with a beautiful singing voice, I can mention this in such an ironic utterance and express my admiration, also ridiculing the other person a little bit in his or her modesty. Sperber and Wilson take the easy way out when they define the exact meaning of irony: To them the ironic meaning is not necessarily the opposite of the literal utterance. There are infinitely many different meanings between a literal interpretation and its opposite, therefore, they deny the classical view that an utterance can either be ironic or not, to them irony is mention and the mention can have all different shades and so can the utterance vary in its ironic strength (Sperber and Wilson, 1981, pg 315).

Is irony as mention still an conversational implicature? Sperber and Wilson (1981, pg 299) assume that conversational implicatures function as common premises that the interlocutors base their dialogue and understanding on. Implicatures as they understand them function as enrichment of the literal sense. As they assume that figurative language like irony communicates something while repressing the literal sense, irony and implicature are incompatible with one another. The one is an addition to the literal meaning, while the other is a substitution for it. In terms of how irony stands compared to implicatures, the relevance theoretic view understands irony as echoic mention. "Instead of figurative meanings, there would be pragmatic implications or implicatures which might carry critical overtones; instead of a failure to distinguish literal from figurative meanings there would be a failure to distinguish use from mention" (Sperber and Wilson, 1981, pg 314). So irony is always a case of mention that conveys some form of criticism. Relevance theory proposes that the figurative sense substitutes the literal sense with the echo of a mentioned utterance. The Neo-Gricean view that we turn next to, understands irony as indirect negation where the graded salience hypothesis predicts that the literal meaning has to be retained and in addition with the non-literal ironic meaning the irony can be derived.

2.3 Neo-Gricean accounts

2.3.1 The graded salience hypothesis

The **graded salience hypothesis** developed by Giora et al. (1998) stands in the middle between a standard pragmatic model and a direct access model, which denies a two-stage process and therefore higher processing cost for non-literal language like irony. According to this point of view, it is salience that determines the activation of processing and not whether something is literal or supported by the context. Thus, highly salient meanings should be processed initially while less salient meanings are processed after the salient interpretation has been activated. For irony the "assumption is that ironic interpretations are less salient than their literal meanings" (Giora et al., 1998, pg 84) and therefore it should take longer to understand irony. Within the graded salience hypothesis irony is understood as a form of indirect negation, called **indirect negation view of irony** (Giora et al., 1998, pg 85), of which the function is to draw attention to a failed expectation. I discuss this approach here among the Neo-Gricean accounts since for non-salient irony it involves **retention** of the primarily activated literal meaning so that the difference between this and the actual ironic meaning can be computed. Even though this stands against Grice and Searle in that they assumed that the literal meaning gets canceled rather than retained (Giora et al., 1998, pg 85), the graded salience theory is in line with these classical theories because it assumes a more complex comprehension process for non-literal language, if it is non-salient. It therefore agrees with Grice and Searle and disagrees with for example relevance theoretic accounts that argue, if the context is rich enough and supports a non-literal meaning, this is accessed immediately and without higher processing cost than literal meaning (Giora et al., 2000, pg 64). But contrasting to Grice, for Giora processing is not affected by literality or context, but salience: the most salient interpretation is always accessed first. In order to be salient, utterances have to be coded in the mental lexicon and additionally they must be prominent due to their conventionality, frequency, familiarity, and prototypicality (Giora et al., 2000, pg 64). That is to say, salient interpretations are assumed to be accessed from the mental lexicon immediately on encountering the linguistic input, whereas non-salient interpretations require extra inferential processes. Then the **retention hypothesis** plays a role. According to this hypothesis, and - as they say - contrasting with Grice, interpreting irony does not involve canceling the literal interpretation and replacing it with the figurative interpretation. Rather, the literal and non-literal interpretation are maintained so that the dissimilarity between them may be computed (Filik and Moxey,

2010, 422).

So the question is now: is irony always non-salient? Because then the predictions this theory makes would be in line with the predictions a Gricean account would make. If there is such a thing as non-salient irony, the predictions of this theory would be similar to those made by relevance theoretic accounts.

The first experiments Giora et al. (1998) conducted were using non-salient irony, so the assumption was that if the irony is less salient it should be more difficult to process and therefore processing should be more slowly.

The two experiments they conducted used reading time in ironically compared to literally biasing contexts. Expected were longer reading times for the ironic contexts. Furthermore they investigated whether in both contexts the literal interpretation would be computed first and that the activated literal concept in case of irony processing is not suppressed but remains active (Giora et al., 1998, pg 85,87,89).

The stimuli were two contexts, one biasing the following sentence towards the ironic meaning, and one towards the literal meaning (Giora et al., 1998, pg 86):

- (11) a. Anna is a great student, but she is very absent-minded. One day while I was well through my lecture, she suddenly showed up in the classroom. I said to her: ...
- b. Anna is a great student and very responsible. One day she called to tell me she did not know when she would be able to show up for my lecture. However, just as I was starting, she entered the classroom. I said to her: ...
- c. 'You are just in time.'

Their results confirm the hypothesis that irony is a two-stage process: "irony comprehension involves the activation of the salient (literal) meaning initially, before activation of the less salient (ironic) meaning. They further demonstrate that the activated literal meaning of irony is not suppressed but remains active as predicted by the indirect negation view of irony" (Giora et al., 1998, pg 91).

In a later study Giora et al. (2000) compared instances of non-salient sarcasm with salient metaphor.

They were also interested in the involvement of the right-hemisphere in the processing of sarcasm. So Giora et al. (2000) conducted a study comparing the performance of right-brain-damaged patients and left-brain-damaged patients in understanding two instances of non-literal language: non-salient sarcasm and salient metaphor.

Since the graded salience hypothesis predicts that sarcasm as form of irony is a two-stage process, where the sarcastic meaning should be derived after and from the salient literal meaning (that is directly retrievable in the mental lexicon), and the right hemisphere (RH) is specialized in linguistic reinterpretation, they expected the right hemisphere to be involved in the comprehension of sarcasm. In contrast, the processing of the salient interpretation should take place mainly in the left hemisphere (LH), where linguistic knowledge in general is stored (Giora et al., 2000, pg 65). They tested left-brain-damaged (LBD) and right-brain-damaged (RBD) patients. Both groups performed worse in the detection of sarcastic meaning compared to the healthy control group. Also RBD individuals performed significantly worse than LBD individuals, when the effect of aphasia was neutralized (that is, language impairment) in the detection of sarcasm (Giora et al., 2000, pg 69). The authors conclude that "there was a selective RH involvement in the comprehension of sarcasm even when the possible contribution of prosody was neutralized" (Giora et al., 2000, pg 70).

There is an issue in the material they used that I think is very interesting (Giora et al., 1998, pg 67):

- (12) Anne and Roger were lawyers in the same firm. Anne hated Roger because he often teased her for defending clients who couldn't afford to pay her fee. One day Anne was at the courthouse while Roger was defending a very wealthy man. He did a terrible job, completely mishandling what should have been a simple case. Anne said to another attorney, "Roger handled that case well."

Questions:

- (13) When Anne said that Roger handled the case well, Anne was:
- a. making a mistake
 - b. telling the truth
 - c. telling a lie
 - d. being sarcastic
- (14) Based on what you heard in the story, which of the following is true?
- a. Roger handled the case poorly.
 - b. Roger did a good job on the case.
 - c. Roger was a tax lawyer.

After the first question they asked their participants one could answer that c. and d. are the case: Anne was telling a lie and being sarcastic. But taking this one step further, in the second question, what is relevant "based on what you heard in the story" as a true answer remains only the first: By

being sarcastic Anne expresses that Roger handled the case poorly. So only when asked about what meaning is the most likely to assimilate in a story context, irony can be detected clearly. Therefore, the most popular incorrect response among those patients that evidently understood the passage, was the "lie" response (Giora et al., 2000, pg 71). Which means that the impaired participants had a **theory of mind**, since they understood that there was some intention in the mind of the speaker, otherwise they would have chosen the "literal truth" response (Giora et al., 2000, pg 71). However, they failed to take the step further that I described with the example.

For metaphors, there was no bias towards the literal interpretation among RBD individuals, while for LBD patients there was such a bias, suggesting that the right hemisphere is more involved in the processing of non-salient sarcasm than salient metaphor. The results support that the left hemisphere is more involved in the understanding of conventional verbal metaphors (Giora et al., 2000, pg 76).

To make sure that understanding was not caused by prosodic cues, they searched for an interaction between hemisphere and understanding and prosody and concluded that understanding in both groups of brain damage is independent of prosody (although the LBD patients scored better on the test when they had prosodic cues, while there was no effect on RBD patients in this regard). They conclude that both hemispheres contribute to the comprehension of sarcasm, but to a different extent, RH contribution seems to be greater (Giora et al., 2000, pg 71–73).

So the question whether the graded salience hypothesis allows such a thing like salient irony or sarcasm remains unanswered. If there exists salient irony then we would expect from the findings just discussed that the left hemisphere plays a more important role in processing that. While for non-salient irony the right hemisphere has proven to be more important.

2.3.2 Relevant inappropriateness

Attardo (2000, pg 816) built his theory exactly on the observation I pointed out to with the example in (12): An ironic meaning evolves because something is either a lie and /or inappropriate to its context but if we consider it as relevant to the discourse we reinterpret it as ironic. He defines two features that have to hold for an utterance so it can be understood as ironic: It has to be inappropriate and relevant to the context, at the same time. Therefore his theory is in favor of a two-stage process, since a seemingly inappropriate utterance has to be reconsidered and then rated as relevant (Attardo, 2000, pg 822–823). He understands his theory of irony as **relevant inappropriateness** as an extension of Grice's account, where irony is a figure of speech,

and not in line with mention theories (Attardo, 2000, pg 794). He has two main criticisms about Grice. One, not only flouting the maxim of quality can derive irony, every other maxim works as well. Two, the firstly derived literal meaning isn't canceled but retains for irony (Attardo, 2000, pg 799, 800). Against relevance theory he argues that the status of an utterance as a mention of something else can only be determined inferentially. How can one detect the mention? Because the statement is inappropriate without. This is why inappropriateness is a necessary and sufficient condition of irony, while mention isn't, since not all cases of irony are cases of echoic mention (Attardo, 2000, pg 806).

Attardo (2000, pg 813) therefore states, the ironic meaning can only be inferred to. That means, furthermore that it is entirely pragmatic and dependent on the context. Third, the process of interpreting the ironic meaning is guided by the cooperative principle (CP) of Grice and, therefore, the CP has to be restored after violation. This last step he calls **principle of least disruption**. Here the "set of shared presuppositions" (Attardo, 2000, pg 814) of Grice play a role: Both interlocutors share the common knowledge that based on the present context the speaker cannot mean literally what he or she said. And so the hearer searches for the implicated meaning. This means the CP has been violated, but after the hearer detected the violation of at least one maxim (and the speaker was counting on him or her detecting that) the CP gets restored. This is what Attardo means with the "smallest possible disruption" (Attardo, 2000, pg 814). That is like communicating to your counterpart: *Hey, I'm going to violate the CP, but only for one conversational unit, so stay with me, it is relevant and will make sense.*

He adds to that, the imperative to be contextually appropriate, which he distinguishes from the need to be relevant. In doing so, he extends the Gricean CP. He defines: "an utterance *u* is contextually appropriate iff all presuppositions of *u* are identical to or compatible with all the presuppositions of the context *C* in which *u* is uttered" (Attardo, 2000, pg 817, italics in the original). While relevance in the framework of Sperber and Wilson is not truth-sensitive, appropriateness is, because it is dependent on the presupposition of the context, that is, it makes a difference if the utterance itself is truth or false. If one lies as an answer, this might be relevant but inappropriate.

Attardo (2000, pg 819) gives the following example, where the answer being a lie is relevant but inappropriate:

- (15) a. Father: 'Did you eat the chocolate?'
- b. Daughter: 'No.' (her mouth is covered in chocolate)

Therefore relevance is not truth-sensitive but appropriateness is. Some-

thing can be relevant and false because it is a lie, but something appropriate has to be true as well.

Attardo (2000, pg 814) writes, "The fact that irony does not necessarily implicate the opposite or the converse of the literal meaning is important". He is in line with Giora in assuming that what makes irony an argumentative value is the indirect negation (Attardo, 2000, pg 816). So Attardo takes from Grice the cooperative principle and shares with Giora's account the assumption that irony is indirect negation and adds to that the idea of a shared set of presuppositions that is called **common ground** by many theories, I will discuss that at length at the end of the theoretic chapter (2.5).

This might get clearer in illustration of an example of Attardo (2000, pg 815) that is sarcastic:

- (16) It is raining.
 - a. What nice weather.
 - b. I was just kidding, as a matter of fact I love rain.

Attardo intends to prove with this example (16a) that irony is indirect negation since I can follow up with a sentence correcting the negated proposition of the ironic utterance. Therefore irony has to be the negation of the literal utterance.

But can one say:

- (17) The sun is shining.
 - a. What a terrible weather.
 - b. *I was just kidding, as a matter of fact I don't like sunshine.
 - c. I was just kidding, as a matter of fact I like sunshine.

The second example would be ironic praise. Adding *I was kidding, I actually love it* should be redundant or at least deletes the irony. Somehow the withdrawal of the ironic statement about the weather does not work here. In this case, there is no personal attitude conveyed one is merely stating that there is good weather outside in a jocular way and maybe there is no personal attitude here that one can refer back to. While the example in (16a) is sarcastic, the example in (17a) is merely ironic.

Every theory that is interested in verbal irony has to account for the very robust intuition that is shared among many theories (for example, Attardo 2000, Sperber and Wilson 1981, Nakassis and Snedeker 2002) that kind irony is much more rare than sarcasm. Attardo (2000, pg 796) notes that positive irony is more risky because if the irony fails, I said something negative, while for sarcasm if it fails what remains is a positive utterance and there can be not much harm in that.

2.3.3 Criticism

As I hope to make clear in the section about humor in (2.4), it is the fact of surprise and failed expectation that makes irony and jokes funny in the first place. So, considering that I assume we use irony to amuse the hearer there would be no point in using it if there were such a thing as a salient ironic interpretation. So far this is in line with the predictions of the graded salience hypothesis for unfamiliar irony: "the graded salience hypothesis predicts that ironic interpretations should take longer to read than their literal counterparts" (Giora et al., 1998, pg 84). However, as at least the psycholinguistic literature claims (eg. Filik and Moxey, 2010) the graded salience hypothesis allows the possibility that there is such a thing like familiar irony. Filik and Moxey (2010, pg 422) assume that familiar irony would be encoded in the lexicon. This is a claim that is questionable, if one considers the retention hypothesis that states that what constitutes the meaning of irony is the gap between what was expected and what is the case. Therefore irony always makes it necessary to conduct the literal interpretation first and then the intended ironic non-literal interpretation to see how they differ. This will be discussed in more detail with the consideration of psycholinguistic evidence in the chapter on psycholinguistic studies in (3.1).

In discussing Attardo's addition compared to the graded salience approach by Giora, he notes that contextual inappropriateness brings along surprise which makes it also highly salient. He points to research on jokes, where it has been shown that a punch-line needs to be unexpected or inappropriate and is therefore highly informative – and funny (Attardo, 2000, pg 821). (This is the point that will be of interest in the section about humor in (2.4).) Attardo also assumes that irony is detected in a two-stage process (Attardo, 2000, pg 823).

Although the two proponents of a Neo-Gricean approach presented here, the graded salience hypothesis and the relevant inappropriateness account, do not make direct claims about how irony and implicature might differ, one could argue that the arguments they bring against Grice's view distinguishes irony from implicature. First, they assume that Grice's approach is too restricted in assuming that only the violation of quality can produce irony as a conversational implicature (Attardo, 2000, pg 799). One can of course be ironic (but maybe not in the most stereotypical sense) by violating, for example the maxim of relevance.

- (18) A and B have been fighting while C was indifferently sitting next to them.
 - a. A asks C: Do you think she is right in treating me like this?

- b. C responds: Don't you guys think the weather today is exceptionally nice?

Here, C is so obviously not obeying the maxim of relevance by saying something completely unrelated that it is clear that he or she means something different than he or she was saying, I would suggest: *I do not want to get into your fight, leave me out of this*. This could still be called an implicature since to Grice the violation of all maxims can derive conversational implicatures.

Second, Attardo and Giroa differ in their assumptions with Grice when they assume that irony is a case of indirect negation. While Grice assumes that the first, literal meaning is rejected when the second, ironic meaning is detected, to them indirect negation evolves because the literal meaning stays active for processing so that the difference between the two can be used to derive the indirect negation (Attardo, 2000, pg 799-800).

For example:

- (19) I have been ill for the last days, when I meet my flatmate in the kitchen, he says: *You look great today*.

I know that he can't be serious because I know that I look ill. So I conclude he means the opposite, that is *You look terrible*. To come to the conclusion that he wants to tell me that I look sick I use his literal utterance. He indirectly negates that he thinks I am being well.

Most importantly, the difference between indirect and direct negation is that direct negation is subject to implicatures, such as scalar ones, while irony as indirect negation is not subject to those implicatures (Attardo, 2000, pg 800).

Another difference between irony and implicature is what Attardo (2000, pg 816) calls the "argumentative value" of irony. This is something, as I mentioned above already, Grice added to his considerations about irony a few years later, too: "irony is intimately connected with the expression of a feeling, attitude, or evaluation" (Grice, 1978, pg 124) I can express an opinion about the issue I am being ironic about. I can do something with an ironic utterance, like evaluate the issue that I talk about, that doesn't seem possible with implicatures. It rather seems to have to do something with the humorous part of irony, that we turn to now.

2.4 General theory of verbal humor

A joke:

- (20) Two matches are walking in the forest. A hedgehog crosses their way. Surprised says the one match to the other: "I did not know that there is a bus going."

One example of a joke from Attardo and Raskin (1991, pg 295):

- (21) The first thing who strikes a stranger in New York is a big car.

And another one from Attardo and Raskin (1991, pg 307):

- (22) How many Poles are used to screw in a light bulb? Five. One to hold the light bulb and four to turn the table he's standing on.

What is true about all this nonsense, is that indeed a hedgehog could look to some child's eyes (this is a children's joke) like riddled with matches. And *strike* is ambiguous and therefore the wordplay works. Also it is possible to screw a light bulb in by standing on a table and turning the person standing on the table holding the light bulb. It is just very inefficient. Jokes surprise us in that they do not meet our expectations. But a good joke does have a truth at its core, something that is ridiculous and not common sense, but still obeys some logic in that it is possible, or true but unheard of in some sense. Therefore, some jokes, just like irony, force us to do some kind of reinterpretation when the punch line occurs.

This brings something very important for irony with it. Irony employs logic that is taken for granted as a means of communication, one could call the Gricean cooperative principle also just basic logic for conversation. On the one hand, the punch line of some jokes ridicules common sense logic but never in an completely arbitrary sense. They trick us in that not the most likely solution is the case but some long-winded solution that still needs to be logically possible. To be funny there needs to be a truth in the logic that surprises us. On the other hand, the ironist says something that is obviously false but surprises us in being true after all, if we reconstruct the ironic meaning. And irony won't work if I just say some kind of nonsense in a given situation, it must be obviously unfitting to the requirements of the present communication.

So, irony and some jokes have something in common: a surprising logical truth. Or with the words of Attardo and Raskin (1991, pg 307): "a joke must provide a logical or pseudological justification of the absurdity or irreality it postulates".

I, therefore, conclude my overview of linguistic theories on irony with their theory of humor, since I think that is a point fairly underestimated in the proposals mentioned so far. Sperber and Wilson (1981, pg 296) ask why we should use ironic speech at all, if not to communicate attitudes or

states of minds in specific situations. In my opinion part of the answer is quite straightforward: to be funny. It cannot be overestimated, I think, how important it is to us to make others smile. That creates closeness and sympathy and therefore understanding and that is the whole point in communication. And what it is that makes us smile is quite concisely explained by the **general theory of verbal humor**.

The general theory of verbal humor (GTVH) can be seen as development of the **semantic script theory of humor** (SSTH) by Raskin (1985) and the **five-level model of joke representation** by Attardo (1989). With it the authors reacted to the critique that the dichotomy between two scripts is too abstract and too much simplifying the phenomenon to be accountable for the complexity of humor (Brône, 2012, pg 467).

Both the SSTH and the GTVH assume that incongruity is a central phenomenon in humor (Attardo and Raskin, 1991, pg 330). This suits well for irony, as being an inconsistency between what is said and what is meant. The SSTH is an approach centered around the mechanism of scripts. A script is a "chunk of structured semantic information" (Attardo and Raskin, 1991, pg 307). The necessary inconsistency of a joke is formed by a script opposition. The meaning of a joke brings two normally unrelated things together (Attardo and Raskin, 1991, pg 310).

Attardo and Raskin detected the quite obvious that most jokes have a similar form, therefore they tried to describe the ways in which jokes can differ from one another. They call them **parameters of joke difference**. They six knowledge resources that inform a joke are also the parameters of joke difference that can be set (Attardo and Raskin, 1991, pg 294–297).

- Parameter 1: Language. This are all the decisions made in terms of language structures, from phonetic, to morphologic, to syntactic and pragmatic levels. (Attardo and Raskin, 1991, pg 298)
- Parameter 2: Narrative strategy. Here the genre is meant. Should one formulate a joke in the form of a question, a riddle, and so on (Attardo and Raskin, 1991, pg 300)?
- Parameter 3: Target. Jokes pick up weaknesses of individuals, or groups, or things. This is always a stereotypical facilitation of reality and as stereotype insulting, but very well known and like this reveals the laziness of our thinking (Attardo and Raskin, 1991, pg 301).
- Parameter 4: Situation. An activity that includes the participants, objects, instruments and anything else that takes place in the joke (Attardo and Raskin, 1991, pg 303).

- Parameter 5: Logical mechanism. That are the kind of manipulations that leave us puzzled, because language is used to describe a situation and in the end we are surprised because our interpretation is not the right one and has been misguided, by wordplays, or what Attardo and Raskin call figure-ground reversals, or false analogies. An example for a figure-ground reversal from Attardo and Raskin (1991, pg 295):

How many Poles are used to screw in a light bulb? Five. One to hold the light bulb and four to turn the table he's standing on.

And another example for a false analogy (Attardo and Raskin, 1991, 305–306):

A student is failing an oral exam in entomology. Finally, the professor shows him the leg of an insect and asks to identify the owner. The student is unable to do that, and the professor flunks him. As the student is leaving the room the professor realizes that he did not put down the student's name. "What's your name, young man?" he shouts after the student. The student sticks his leg back into the room and says, "You guess it, professor."

They also mention the logical mechanism of "false priming" or "garden path" (Attardo and Raskin, 1991, pg 306), just as in the sentence *An astronomer married a star*.

The default version of a logical mechanism according to Attardo and Raskin, are juxtapositions, where the ambiguity or homonymy of words is exploited in a funny way:

The first thing who strikes a stranger in New York is a big car.

In general, they postulate "a joke must provide a logical or pseudo-logical justification of the absurdity or irreality it postulates" (Attardo and Raskin, 1991, pg 307).

- Parameter 6: Script opposition. A script is evoked as a chunk of semantically related interpretations. The core assumption of SSTH is, that the words of a joke are

[...] always fully or in part compatible with two distinct scripts and that the two scripts are opposed to each other in a special way. In other words, the text of the joke is deliberately ambiguous, at least up to the point, if not to the very end. The punchline triggers the switch from the one script to the other by making the hearer backtrack and realize that a different interpretation was possible from the very beginning (Attardo and Raskin, 1991, pg 308).

Attardo and Raskin (1991) add to the script opposition, which was central to the SSTH, five other knowledge resources that are found in most jokes. They get six knowledge resources (KR) that inform a joke: Script oppositions, logical mechanisms, situations, a target, a narrative strategy and language. "Each KR is a list or set of lists from which choices need to be made for use in the joke" (Attardo and Raskin, 1991, pg 313). Situational jokes, they assume, are made out of these components. They distinguish between content dependent and independent KRs. The principle they formulate is: in order to make a joke, "one would need to observe some of its components in the current situation and to provide all the missing ones" (Attardo and Raskin, 1991, pg 326). One takes the situation and enriches it, kind of, with the features that are missing.

Attardo postulated in 1987 his **five-level model of joke representation**. In this top-down model the lowest level is the text of the joke, the second level is a specific language form of the joke, the third level consists of the target of the joke, the fourth level he called template, that is two opposing scripts and logical mechanisms. The final level combines the script-opposition and the logical mechanism (Attardo and Raskin, 1991, pg 309). Informed from this, they assume a hierarchical ordering.

Table 2.1: the general theory of verbal humor

Attardo's five level hierarchy		Knowledge resources
5	Basic	Script opposition
4	Template	Logical mechanism
3	Target + Situation	Situation
		Target
2	Language	Narrative strategy
1	Surface	Language

(Attardo and Raskin, 1991, pg 309, 311)

As this is a hierarchic model, the upper levels determine the lower ones. (Attardo and Raskin, 1991, pg 311) But, "a lower level is not a later level" (Attardo and Raskin, 1991, pg 327). They do not assume that comprehension of jokes actually takes place in this order, but try to analyze the components of jokes in an analytic way. Attardo and Raskin stress that this should not be confused (Attardo and Raskin, 1991, pg 327–328).

Criticism

Of course, one can criticize their theory, as Brône (2012, pg 470) does, for being too abstract. But since they themselves stress that they formulated the theory in order to provide a formal typology to categorize jokes their theory should merely be seen as a first step towards understanding how jokes work. The difference between abstract theorizing and actual processing will be of importance again when talking about psycho- and neurolinguistic studies in (3).

2.5 Presuppositions

So far one of the aims why I presented different theoretical approaches to how the meaning of irony can be described was to find out how irony and implicature are different and where these competing theories draw the line between implicature and irony. What remains now is to take a look on how presupposition and irony play a role for each other.

We will first have a closer look on what presuppositions are, since this leads to the notion of common ground that is central to pragmatics and cannot be left out by any attempt to understand irony.

Russell (1905) challenged the theory of Frege (1948) in his paper "On Denoting" in that he had counterarguments against the distinction between **meaning** (**sense** in Frege's framework) and **denotation** (the **reference**, as Frege called it). The following, as a case where denotation is absent. The very famous example from Russell (1905, pg 485):

(23) The present king of France is bald.

The king of France compared to the sentence *the king of England* (at a time when there existed a king of England), the former has a meaning but no denotation, while the latter has both (Russell, 1905, pg 483). The sentence cannot be true, because there is no king of France. As Russell states: "Hence one would suppose that 'the King of France is bald' ought to be nonsense; but it is not nonsense, since it is plainly false" (Russell, 1905, pg 484). So Russell observed that only some meanings have denotations (Russell, 1905, pg 487).

There are two kinds of conditions that have to be satisfied for an utterance to be felicitous: existence and uniqueness. This means there must exist a king of France and only one, that we can refer to unambiguously. This are prerequisites for a sentence to have a truth value. Such a prerequisite is also

called **presupposition**.¹ Therefore we have an existence presupposition and an uniqueness presupposition that have to be the case that there is an extension for *the present king of France* (Zimmermann and Sternefeld, 2013, pg 210). This is about entailment, because to utter the sentence *The present king of France is bald* entails that there is a king of France (existence presupposition) and that there is only one (uniqueness presupposition). According to the Fregean analysis of this, the entailment holds even for the negation:

(24) The present king of France is not bald.

meaning, this also entails the existence of exactly one king of France (Zimmermann and Sternefeld, 2013, pg 211). It is one of the defining characteristics of presuppositions that they are not blocked by negation (Zimmermann and Sternefeld, 2013, pg 228).

For verbs like *know*, there is an entailment in the sort of: if I know that something is true, one can infer that this something must be true for the actual world. The **inference** is lexically triggered, that means it is part of the lexical meaning of the verb, that the inference is permissible.

"[I]f x knows that p in w , then p is true of w ." (Zimmermann and Sternefeld, 2013, pg 212)

This holds also for the negation *didn't know*. If someone didn't know something was true, this must have been true.

"[I]f x does not know that S in w , then S is true of w ." (Zimmermann and Sternefeld, 2013, pg 212)

The two taken together: "S must be true if any world in which x knows S, and that S must be also true if any world in which S does not know that S" (Zimmermann and Sternefeld, 2013, pg 212). We have a tautology if a sentence and its negation are always true in logical space. The sentence *S or not S* is such a tautology because the truth value is always 1, since it is always true (Zimmermann and Sternefeld, 2013, pg 159). Now the definition of *know* makes S sound like a tautology, but the sentence *I know* clearly isn't always true, so no tautology possible. As it was the case for descriptions, that they presuppose existence and uniqueness of their referent, we now have to permit that a verb like *know* presupposes the truth of its complement. Now the sentence *I know something* can be false, if the *something* has no truth value but a gap. So the extension of the sentence is neither false nor true,

¹Frege called it "Voraussetzung", prerequisite, when he talked about existential presuppositions: "Wenn man etwas behauptet, so ist immer die Voraussetzung selbstverständlich, daß die gebrauchten einfachen oder zusammengesetzten Eigennamen eine Bedeutung haben." If you say something, the prerequisite is natural that the simple or composed proper nouns have a meaning (Frege, 1892, pg 40, my own translation).

but empty. That is called **truth value gaps** (Zimmermann and Sternefeld, 2013, pg 212–214).

A presupposition, now, is defined, as follows:

If a sentence S_1 entails a sentence S_2 , and if the negated sentence 'not S_1 ' also entails S_2 , and if S_2 is not a tautology, then S_2 is called a **presupposition** of S_1 (Zimmermann and Sternefeld, 2013, pg 215, boldface in the original).

We assume now that presuppositions are like preconditions for any truth value of a sentence. The presuppositions of a sentence S split the logical space in more than two parts, true and false, but also no truth value if the preconditions are not given. The part where S is true is called **positive intension**. The part where S is false, is called **negative intension**. The part in logical space where S has no truth value is the remaining set of worlds. In the latter case, the function from worlds to truth values, that intensions are, is undefined for these worlds. This is called "undefinedness" or "partiality" (Zimmermann and Sternefeld, 2013, pg 216). It turns out, that this is a feature quite widespread in natural language that takes place not only in the meaning of sentences, but also lexical items. Words have conditions that have to be accounted for, in order to use them in a sentence that makes sense. Therefore presuppositions can be triggered by lexical items alone. I use the example that Zimmermann and Sternefeld (2013, pg 217) use because it is related to the famous example by Russell: The predicate *bald* is true for the set of people that are bald, it is false for the set of people in the actual world that have a lot of hair, meaning they are not bald, and it does not make sense for the remaining people who do not have a head in the first place. Truth conditions therefore have to be analyzed as positive and negative extensions, in accordance to positive and negative intensions (Zimmermann and Sternefeld, 2013, pg 215–217).

As was mentioned in the precedent paragraph, presuppositions can be **triggered by lexical items** alone. Zimmermann and Sternefeld (2013, pg 231) argue, thus, that presuppositions are part of the literal meaning of a sentence, and therefore in the domain of semantics. What we have seen so far is how presuppositions are explained in a theory that understands them to be a part of semantics.

The other point of view are theories that treat presuppositions to be in the realm of pragmatics. I would like to review the very influential approach by Stalnaker (1973) as a proponent of a theory that understands presuppositions to be pragmatic phenomena. He understands this to be in contrast with the standard notion of presuppositions, defined as one sentence presupposing another when the following sentence must be true for the former to

yield to any truth value whatsoever. He understands propositions as relations between what a person says and what he or she takes for common knowledge, instead of assuming as a relation between propositions and sentences themselves (Stalnaker, 1973, pg 447).

Statements and requests are made, questions asked, proclamations and commands issued, against a background of common knowledge, or at least what is represented as common knowledge. This background of knowledge or beliefs purportedly shared by the speaker and his audience constitute the presuppositions which define the context. A rough definition might go something like this: *A speaker presupposes that P at a given moment in a conversation just in case he is disposed to act, in his linguistic behavior, as if he takes the truth of P for granted, and as if he assumes that his audience recognizes that he is doing so* (Stalnaker, 1973, pg 448, italics in the original).

For Stalnaker then, a presupposition is a propositional attitude of accepting something to be true. These attitudes make up the set of possible worlds, that is, the set of worlds that the presupposition is true for. He calls that the "presupposition set". Also, what comes with this definition of presuppositions is the assumption of "general principles of rational behavior" (Stalnaker, 1973, pg 450) from which maxims can be derived. This sounds very much like the conversational maxims of Grice and the conclusions Stalnaker draws from it, are also similar: As soon as such rules are present, the possibility of "exploiting" (Stalnaker, 1973, pg 451) them by pretending shared beliefs, is present, as well. He even considers this possible pretense as a way of communicating. Which reminds us of the view that irony is pretense, that was mentioned in the chapter on theories of irony. Presuppositions even work, when common knowledge is only pretended, therefore, presuppositions can't be equal to what is commonly believed. He distinguishes the cases where a person can choose to make a proposition, from the instances where a sentence requires a proposition. Presupposition requirement is the case if a sentence needs to presuppose something because it would be weird, otherwise. These cases are for Stalnaker the instances where the linguists' narrow and semantic notion of presuppositions being relations among sentences applies. What from the semantic point of view is called presupposition, he calls in his pragmatic account presupposition requirement. But to limit the understanding of presuppositions to the technical understanding of the purely semantic account does not do justice to the various sources of presupposition requirements (Stalnaker, 1973, pg 451–452).

The semantic account of presupposition understands them to influence the truth value of the sentence depending on the presupposition that is false or true. So Stalnaker presents arguments to reject this semantic claim, in that they show the truth or untruth of a sentence to be independent of the truth or untruth of its presupposition: One argument involves the word *even* (Stalnaker, 1973, pg 453).

(25) The weather might even be nice. vs. The weather might be nice.

For the two sentences what is asserted stays the same, with or without the *even*. That is, there is the possibility that the weather will be nice. What *even* presupposes is that there are also other possibilities and that it would be somewhat surprising if the weather would be nice. But the truth value of the assertion is unrelated to the presupposition.

Another argument: If the semantic notion of presuppositions describes required presuppositions, a pragmatic approach that claims to explain the phenomenon in a more general perspective must also account for other more general constraints that evoke presuppositions. One such constraint is that there are sentences that exclude certain presuppositions and are weird if they presuppose something that does not fit to the sentence and the context. Another constraint is that there are sentences that have to presuppose a certain kind of presupposition, but do not require a specific one out of this set. Like demonstratives in his example: "She is a linguist" (Stalnaker, 1973, pg 454). Requires that there is one female person that is a linguist. But in different contexts there can be different women that this description matches to. It is just a set of referents but not a certain referent that is needed for this presupposition.

His last argument for a pragmatic notion of presuppositions concerns the projection problem that I will come at length to at the end of this chapter in (2.6.2). At this point only so much: This problem deals with the fact that compositionality does not seem to apply for sentences with presuppositions if it is possible that one of the components of a complex sentence presupposes something that the entire complex sentence lacks to presuppose. If one takes the pragmatic point of view, however, one can just assume that background assumptions change while a dialogue unfolds. Presuppositions can just be "denied, challenged, retracted or forgotten" (Stalnaker, 1973, pg 455) over time. So it is not a problem that a phrase seems to presuppose something that does not hold anymore if the phrase is put in another surrounding context.

In general, Stalnaker (1973) suggests to separate truth values as semantic term from presuppositions that as pragmatic phenomena would not interfere with the truth value of a sentence.

2.5.1 Common ground

Presuppositions as part of pragmatics require a shared background of common understanding on the part of both interlocutors. This is called the **common ground**. It has to be present in order to understand the proposition as a set of possible worlds by being able to exclude some states of affairs and permit others, finally extracting the new information of a sentence. Therefore, **new information** is relative to the common ground: what is part of the common ground cannot be new information, the presupposition has to be part of the common ground. Zimmermann and Sternefeld (2013, pg 222) define the common ground as "the set of possible worlds that are compatible with [what is assumed to be] the discourse participants' shared assumptions at a certain point in conversation". **Updating** is what changes the assumptions of the interlocutors during discourse by excluding some worlds of which S is false and narrowing down the ones of which S is true. Beginning with the discourse we have the common ground CG, and with updating we get CG', which is what remains from CG after the utterance of S. $CG - CG'$ is the **context change potential** of a sentence S. If the presupposition p is entirely included by CG then the sentence S is completely uninformative. The proposition minus the presupposition is the information one derives out of a sentence (Zimmermann and Sternefeld, 2013, pg 222–224).

It is not entirely true, though, that the proposition of a sentence does not entail new information for the interlocutor. On the contrary, we can and do communicate information while presupposing it in a very efficient way: I can tell someone *I thought it would not be good weather today* presupposing that there is good weather even if the person that I am talking to does not know which weather we have, for example because we are on the phone and my addressee is in another city. The person will understand two bits of new information that I have given: a) we have good weather here (that is what I presupposed) and b) that I did not expect that kind of weather (what I asserted straight away). What happens is that the presupposition is precisely not part of the CG here. This is made to work by the hearer that adjusts his CG according to the sentence he or she has heard. This is called **accommodation**. I can expect from my opposite to put a presupposed proposition into its common ground that is not there prior to the utterance but he or she has to do that before the update of new information (Zimmermann and Sternefeld, 2013, pg 226, 227).

This is a crucial difference between irony and presupposition. I cannot make an ironic comment on something that is not already in the common ground of speaker and hearer. If I say ironically on the phone *We have such great weather here* and the person at the other end of the line is in another

city and does not know that it has been raining here for days, the irony will fail.

Different notions of common ground

Since this is quite central for presupposition in pragmatic terms and for the sake of completeness, I will revise some of the most recent theories that deal with common ground and updating-operations. This is supposed to show that this is a much discussed topic in the literature on discursive speech.

Commitment change potentials Beyssade and Marandin (2006) propose a framework based on the observation that natural language shows that the same utterance can be used to convey different kinds of speech acts and also may be used to perform more than one speech act simultaneously.

Clause types are defined by two features: Semantic ones and combinatorial ones. Semantic features are described by a CAUSALITY dimension, while combinatorial features are connected to grammatical functions by the dimension HEADEDNESS. The HEADEDNESS dimension defines the possible syntactic combinations, while the CAUSALITY dimension specifies the semantic constraints that identify a certain type of content (CONT). There are four subtypes of clauses (declaratives, interrogatives, imperatives, and exclamatives): (Beyssade and Marandin, 2006, pg 45,46)

- (26) a. *decl* – *cl* → [CONT *Proposition*]
- b. *inter* – *cl* → [CONT *Question*]
- c. *imp* – *cl* → [CONT *Outcome*]
- d. *excl* – *cl* → [CONT *Fact*]

Illocutionary forces are understood as "types of moves" that come with different "commitment change potentials" (Beyssade and Marandin, 2006, pg 42,43). From this perspective, conversation is understood as conversational game where the act of communication is understood as intended and it is signaled by the addressee that the intention has gone through. Irony is like one more advanced play in this game that comes with different effects on commitment. These commitments take place on the speaker's level first. The speaker commits to something and then wants the addressee to change his or her commitments accordingly. The crucial point for irony is that these commitments do not have to be proportional to each other. Also there is what speaker wants from the addressee, this request is called the call on the addressee. These different intentions of a discourse find expression in discourse by updates. So what utterances do in conversation is a constant

process of updating, of adaptation relative to meaning, commitment and requests. What is important is the updating or adaptation process that allows the interlocutors to decide what clause type comes with which illocutionary force. The special type of update the speaker wants the addressee to perform is called speech act assignment. It might be marked by

- (27) a. contextual means (inferences)
 b. grammatical means (lexical markers, constructional markers, prosodic markers)

Contextual means that belong to the clause type alter speaker's commitment, grammatical means that do not belong to the clause type mark speaker's call on the addressee. The question is, thus, who has which information available to coordinate the update with.

Speech acts are concrete implementations of illocutionary forces. Therefore the speaker can commit to the content of the four clause types above. He or she can commit to a proposition, a question, an outcome, and a fact. (The latest is a special case, as we will see.)

They get a dialogueic gameboard (DGB), as they call it, like this:

(28)

$$\left[\begin{array}{c} DGB \left[\begin{array}{c} SG \\ QUD \\ TDL \left[\begin{array}{c} SPKR \\ ADDR \end{array} \right] \\ CALL - ON - ADDRESSEE \end{array} \right] \end{array} \right]$$

Uttering an assertion that corresponds to a set of homogeneous propositions updates shared ground (SG). Uttering a question updates question under discussion (QUD). Uttering a directive updates to-do-list (TDL) by speaker or addressee. The public part of DGB that is shown here gets updated only when both dialogue participants (DPs) have accepted the update of the proposition in discussion. CALL-ON-ADDRESSEE relates to the feedback that the speaker demands. But in contrast to the three move types described so far, the commitment to facts behaves differently. Exclamatives do not require a commitment of addressee.

What we get are commitments that require both interlocutors to update and commitments (CMT) that are only performed by speaker's update:

(29)

$$\left[\begin{array}{c} DGB \left[\begin{array}{c} INTERACTIVE - CMT \left[\begin{array}{c} SG \\ QUD \\ TDL \end{array} \right] \\ SPEAKER - ONLY - CMT \\ CALL - ON - ADDRESSEE \end{array} \right] \end{array} \right]$$

The distinction they draw between simple and complex speech acts is that for complex speech acts speaker's commitment and speaker's call-on-the-addressee are distinct while they are identical in simple speech acts. Interestingly they conclude that there are speech acts for that the type of call-on-the-addressee has to be inferred by what the addressee knows about the believes to conclude how the content of speaker's commitment is different from the call-on-the-addressee.

Tabeling Malamud and Stephenson (2015) will help us to answer the following questions: What happens to what the speaker wants the addressee to do with his commitments and what to the call-on-the-addressee by the speaker? Their first modification is that they assume projected commitments. They thus propose a conversational scoreboard with the following elements to account for the semantics and pragmatics of dialogueue (Malamud and Stephenson, 2015, pg 9):

- (30) a. DC_x : participant X's public discourse commitments
- b. Table: issues of the conversation represented as propositions where the latest and top issue is also on top of the stack; issues can pend on the Table, i.e. being part of the Common Ground, when they aren't resolved yet
- c. CG: the Common Ground is the set of propositions that all speakers are publicly committed to
- d. Projected CGs: potential CGs that might help to resolve the issue on the table, that is a set of propositions that all participants are committed to, plus possible resolutions to the issue on the table; in other words what the participants assume to be the next stage of conversation

These elements are simultaneously part of the conversational scoreboard and are also included in a three step sequence: *previously*, *after A's assertion* and *after B accepts A's assertion*. As an example, the following question:

- (31) A and B have been to the movies. A asks: *That was a good movie, right?*

Table 2.2: the conversational scoreboard

	previously	after A asks	after B answers	after A accepts B's answer
DC _A	{}		{}	{}
Proj.DC _A				
DC _B	{}	{}	{The movie was good}	{}
Proj.DC _B				
Table	<>	<{The movie was good, The movie was bad} >	<{The movie was good} >	<>
CG	{q}	{q}	{q}	{q, The movie was good}
Proj.CG _s	{{q}}	{{q, The movie was good}, {q, The movie was bad}}	{q, The movie was good}	{q, The movie was good}

(see Malamud and Stephenson, 2015, pg 15)

Note also that there is a distinct set of common standards (CS) they assume to be a different part of the scoreboard. It includes the standards that have been agreed upon for the purpose of the conversation. As one can see in this paragraph's table, the projected set of CG is the table plus CG. In the CG are only the propositions that all participants are committed to. What they add, furthermore, to account for the fine grained modifications of the dialogueue that discourse markers are capable to produce, is projected commitments for all the participants. These projected commitments of the speaker and the hearer(s) are what they expect to be the next stage of conversation. Speaker's commitments are a special case, though. Since the Speaker is under control of the dialogueue right now he or she can decide to make a projected commitment rather than a present one. If that is the case, the hearer(s) can infer that the speaker has a reason to delay his or her commitment, as he or she does by making a projected commitment. "Pro-

jected commitments give rise to an implicature of tentativeness” (Malamud and Stephenson, 2015, pg 14).

Another modification Malamud and Stephenson (2015) make is that they account for metalinguistic issues in their analysis. They propose that some speech acts do not only make an ordinary conversational move, but also raise metalinguistic issues of that move. The assumed tentativeness of moves, that was just mentioned, requires updates of projected commitments rather than present commitments. Moves that involve commitments of that kind and raise a metalinguistic issue are pending the resolution of the metalinguistic issue (MLI) as well. ”MLI_{*p*} is a contextually determined set of propositions, any of which would resolve the contextually determined metalinguistic issue concerning *p*” (Malamud and Stephenson, 2015, pg 20). This MLI is put onto the stack of the table after *p* itself and is added to the projected CGs, that means MLI has to be resolved by the addressee before the proposition itself can be updated. Irony could be such a metalinguistic issue that has to be detected first, before the dialogue can move on in this direction. Otherwise the irony fails and alternatively the literal interpretation is updated instead.

Grounding Clark and Brennan (1991) define grounding as process that guarantees the updating of common ground as necessary condition for communication. When something becomes part of the common ground it is agreed by the participants of the conversation that they reached mutual understanding to a criterion sufficient for present purposes. They call that ”grounding criterion” (Clark and Brennan, 1991, pg 129). In fact people do not only seek the absolute level of grounding but positive evidence as a higher criterion. This positive evidence can come in acknowledgments (like, back-channel responses), the initiation of the next turn (formed in adjacent pairs, this observation is based on the assumption that conversation generally follows standardized coherent sections with entries, bodies, and exits) or continued attention. For irony it is proposed, as psycholinguistic literature will show (see Gibbs, 2000) in the next chapter in (3), that an appropriate response might be an ironic answer. For Clark and Brennan grounding takes place in two phases: presentation and acceptance. After the hearer accepted the contribution to the dialogue it is complete (Clark and Brennan, 1991, pg 130). But after what we have seen so far, it takes more to successfully update the common ground.

Bavelas et al. (2012) account for this in that they assume a three step process. They expand traditional models as the one by Clark and Brennan most importantly in one way. The speaker’s response to the addressees feedback is not only that he goes on and keeps on talking but an observable reaction.

There is a third layer of grounding that makes the speaker's acknowledgment of the addressee's response explicit. Thus the minimum grounding sequence, and therefore unit of analysis for dialogue, is:

1. Speaker presents information.
2. Addressee shows his understanding.
3. Speaker gives feedback.

As we have seen above, utterances in this sense only form a dialogue in their functional relationship to each other. This grounding sequence is by definition very sensitive to changes. The last change decides on what the interlocutors ground on. This reminds one of the table as a stack where the latest move gets on top and has to be resolved first. With grounding it works analogous and according to Bavelas et al. (2012) it is these changes to the original presentation that happen in natural discourse that make the three step as default necessary, and also might be able to account for sensitive agreement between the interlocutors, such as it is the case for indirect speech.

Problem Those theories on common ground are to be understood as attempts to grasp the complexity of discourse in natural language. Even though there are attempts to describe implicatures or metalinguistic aspects in the framework, I do not discuss these issues further. I leave it here at that point, because the theories themselves are a matter of debate and thus not sufficiently formulated to apply them already to cases like irony.

2.6 Discussion

2.6.1 Implicatures and irony

There is no doubt that implicatures can be used to be ironic. But what is the relationship between them exactly? This will be the central question of this discussion. Grice's own example for a conversational implicature in (2) shows that irony cannot only be derived by violation of the maxim of quality. The example is repeated here for convenience (Grice, 1967, pg 43):

- (32) A asks B how C is getting on in his job, and B replies, *Oh quite well, I think; he likes his colleagues, and he hasn't been to prison yet.*

Here the speaker B implies that the most one can say about C's new job is that it hasn't brought him in prison yet. He does not say something that is untrue and is therefore obeying the maxim of quality. I would argue

that this is still irony because he says something that does not seem relevant to the dialogue if it is understood literally and it is funny. The question was how C is doing in his new job and B tells as an answer that he hasn't been arrested yet. If C is not a professional criminal, in this case not being arrested would be a direct credential for his capability, but this is not a likely possibility, so it is a joke. The question is now what is the implicature here and what the irony? Assuming that the cooperative principle is intact, we conclude that this is the most relevant and the most true thing B can say. Which implies that his colleagues might be treacherous people, or C might be a person easily tempted to do something against the law given his new occupation, as are possible implicatures that Grice points out to. What is funny is the understatement, with which B is implying that the most he can say about C's job achievements is that he hasn't done something criminal or hasn't been caught at least. And this is of course something that you should achieve in your job, but it is the very least you should be able to do, therefore, it is funny to utter this as a response. The implication is a judgment about C and his coworkers, the irony is that B chose to answer the very least one can say about someone doing his job reminding us that everyone should at least remain in accordance with the law while working. He makes a joke about C at the same time, because all he can say about his new work isn't much of an achievement for C. There would be no irony if the new position of C would be of the kind where the fear that he might not stick to the law is justified, because it is a preoccupation that likely could lead to someone being arrested. Then it would be too close to reality to be funny and not evoking an alternative background. The implicature about his coworkers being treacherous people of C being a person easily tempted could still be there, though. To tear apart the different possibilities, whether C is a person easy to seduce, or whether his new job is dangerous, or whether B is just making a joke, a great deal of information about C or his new job is necessary. Interestingly, the less likely the two former possibilities are, the funnier the utterance would be. Irony and implicature stand in a relationship with each other, the higher the possibility that B wants to make a comment about C's coworkers, the smaller the possibility that he wants to make a joke. And the less likely C's coworkers are mean people, the funnier B's comment is, because the contrast is higher. And as we saw with jokes, it is about inconsistency. That means, the less strong the conversational implicature, the stronger the irony. So there seems to be the tendency that Potts observed for conventional implicatures, that are either implying something or being ironic: They are in a continuum with each other, in that the one can be derived by the other but the strength of one comes at the cost of the strength of the other.

To make this clear, another example

- (33) a. A: She is smart and beautiful.
 b. B: Yes, she is smart.

What B implies here by not obeying the maxim of quantity, is that he should agree to all the points A made about another person, but B chose to omit something. Which implies that he does so on purpose and wants to communicate with it that he or she does not agree that she is both, smart and beautiful. The implicature is: To B she is only smart. This is close to scalar implicatures that will be the issue of discussion in just a bit. With scalar implicatures we assume that there is an implicit range, and by stating something it is implied that the logically stronger possibilities are excluded. Saying someone is smart and beautiful is a logically stronger claim because it rules out more possibilities than saying someone is smart or someone is beautiful. This (scalar) implicature can be used ironically in an appropriate context. For example, C has done something stupid and A does not know that and says *She is smart and beautiful* and B who knows that C did something stupid answers *Yes, she is beautiful* being clearly ironic and meaning that she is not smart. The classic case of irony would be for B to say *Yes, she is smart* and meaning the opposite. This shows that irony can also be derived by omitting salient, recent information.

In the clear cut example for irony of the fine friend, one could assume that the irony and the implicature are the same thing. There is nothing there besides the float of the maxim of quality. He just says obviously the opposite of what is true. So irony and conversational implicature can be the same, they don't have to be, though. Irony can also be derived by other means like understatement, exaggeration or the obvious omission of information. This shows that Grice's account does not go far enough in some respects.

Giora et al. (1998, pg 85) argued against Grice because his account assumes that the literal meaning gets suppressed and rejected. However, for Giora and other Neo-Gricean accounts irony is an enrichment of the literal meaning that is computed (at least for non-salient irony) in two steps. So both the ironic and the literal meaning have to stay accessible. According to Sperber and Wilson (1981) Grice's conversational implicatures are an addition to the literal sense, in that they function as "premises in an argument designed to establish that the speaker has observed the maxims of conversation in saying what he said" (Sperber and Wilson, 1981, pg 299).

What does Grice actually say about the issue whether the literal interpretation remains to compute the irony? Only so much:

Since, to calculate a conversational implicature is to calculate what has to be supposed in order to preserve the supposition

that the Cooperative Principle is being observed, and since there may be various possible specific explanations, a list of which may be open, the conversational implicature in such cases will be disjunction of such specific explanations; and if the list of these is open, the implicature will have just the kind of indeterminacy that many actual implicata do in fact seem to possess (Grice, 1967, pg 58).

This is an interesting quote, since a implicature being the disjunction between a list of possible meanings sounds like literal and non-literal meaning have to be sustained to compute the implicature, which is exactly what Neo-Gricean approaches propose. Grice (1967) writes for conventional implicatures that something "follows" from using an implicature. For conversational implicatures Grice merely says that a conversational implicature has to be "worked out" while a conventional implicature can be "intuitively grasped" (Grice, 1967, pg 50).

According to Grice conversational implicatures, as part of pragmatics, can be deleted while assertions and conventional implicatures as part of the semantic meaning cannot be deleted.

While for (conversational) implicatures in the Gricean sense the case is that the speaker means proposition p and the relationship between literal and non-literal meaning is: the latter is a conclusion of the former. For irony on the other hand, as Attardo assumes the speaker wants to negate p . That means, there is a basic difference between implicatures and irony: in the case of implicatures the literal interpretation is enriched and specified while for irony the literal interpretation should be replaced by its negation. However, I intend to make some arguments that irony functions as enrichment of meaning, too.

Implicatures function as enrichment of meaning since they strengthen the logic implications in that they enrich the meaning towards the logically stronger meaning (that is the one that excludes more alternatives). The implicature for the example (34) below would be that I have *exactly* two Euro. But the implicature gets canceled with b. in favor of the logically weaker *at least* two Euro-reading. Not only do I have two I do have even five Euro, this is logically possible since five entails two. That is why we speak of a scalar implicature and they can be canceled (Panizza et al., 2009, pg 6):

- (34) Do you have two Euro that you can borrow me for a cup of coffee.
 - a. Yes, I have two Euro.
 - b. In fact, I even have five, so you can take a cake to your coffee.

The view that the strengthening of meaning is derived by a rather complicated process through Gricean reasoning is contested by other theorists (for example, Chierchia, 2013). They assume that this mechanism, and its underlying logic, which says that strengthening, in other words, making a scalar implicature, works in an upward entailing context, but not in a downward entailing environment, is in fact part of the grammar (Panizza et al., 2009, pg 6). Irony, on the other hand, should negate the straightforward meaning. An interesting observation can be made with the example of Attardo (2000, pg 815), that I would like to repeat here, where he wants to show that irony is indirect negation.

- (35) It is raining.
 a. What nice weather.
 b. I was just kidding, as a matter of fact, I love rain.

But what can be seen here, as well, is that irony can be canceled.

Another example in which the irony gets deleted (at least partially) is the following:

- (36) The princess grabs Luke's gun and fires at a small grate in the wall next to Han, almost frying him.
 HAN: What the hell are you doing?
 LEIA: Somebody has to save our skins. Into the garbage chute, wise guy.
 She jumps through the narrow opening as Han and Chewbacca look on in amazement. Chewbacca sniffs the garbage chute and says something.
 HAN: Get in there you big furry oaf! I don't care what you smell! Get in there and don't worry about it.
 Han gives him a kick and the Wookiee disappears into the tiny opening. Luke and Han continue firing as they work their way toward the opening.
 HAN: Wonderful girl! Either I'm going to kill her or I'm beginning to like her. Get in there!

(Lucas, 1977)

In this dialogue we have two instances of irony. The first one is a clear cut example for irony: By saying *Wise guy* Princess Leia clearly means the opposite, namely, Han Solo not being wise at all, but stupid. On the other hand, Han Solo saying *Wonderful girl*, judging by the dialogue they had before is definitely ironic, more precisely sarcastic, but the ironic meaning

partly gets deleted by him saying *or I'm beginning to like her*. It deletes the opposite meaning and requires the literal interpretation to make sense, too.

Having shown that, I think that an ironic meaning can even be left ambiguous. Utterances can be left undecided as an offer to be understood ironically or not. Not very strong and rather subtle kinds of ironic utterances can function like interpretation-offers. The hearer can choose whether he or she wants to pick the literal or the ironic interpretation. And the speaker can provide an additional ironic layer for those hearers that have enough knowledge to detect the irony. Those who know the speaker well enough - because they are aware of the situational context or have information that other interlocutors do not have - can compute the ironic meaning. I think it is possible that a slightly ironic utterance can lead to different interpretations among the hearers. Irony, in the end, has to be construed by the opponent, who decides whether an utterance is ironically or literally meant. As Zhao (2011, pg 180) quite poetically says: "Verbal irony, like beauty, is in the eye of the beholder."

Potts (2005, pg 28) observed for conventional implicatures that an utterance can either have a conventional implicature as meaning or an ironic meaning, but only one or the other. Attardo (2000, pg 800) claimed that irony as indirect negation is not subject to scalar implicatures that rule direct negation. Taking this together with Potts observation, it might be the case, since scalar implicatures optimize informativeness, that irony optimizes informativeness in its own right and therefore enriches logical mechanisms, just like scalar implicatures do.

The hypothesis is that irony works like scalar implicatures in that if computed they both strengthen the logical implications in that they function as enrichment of meaning. If irony is an enrichment of meaning like scalar implicatures that strengthens the logical implications of an utterance this can also explain why kind irony is much more seldom than sarcasm. Just as Attardo (2000, pg 796) pointed out, positive irony is more risky because if the irony fails only the literal criticism remains. While for sarcasm the critique is hidden behind a positive literal meaning. And it is also a riddle why we use irony so often, even though the risk of failure is so big. I am not aware of a study that shows how often irony is not understood, but I assume it must be very likely given the very sensitive clues that one needs to detect (something that is the issue in the chapter about psycho- and neurolinguistic studies in (3)). That this huge possibility that the ironic meaning is not detected is on the one hand a hint on how much we apparently enjoy it, so that we take the big risk. On the other hand, assuming irony is a conversational offer, an additional layer of meaning for those who are adept enough, might explain why we take the risk of being misunderstood so freely. And as we

saw with jokes, they function as enrichment of meaning, too. What Attardo and Raskin (1991, pg 326) describe is that for a situational joke you need to observe all the present cues in a certain situation and add the missing clues to derive a meaningful description out of the situation. So one enriches meaning with the knowledge resources they describe. The same is true to understand a joke, you have to add meaning in accordance to the situation.

The assumptions made are clearly in favor of accounts that assume the literal meaning to remain active after the non-literal ironic meaning is activated, namely the graded salience and the relevant inappropriateness account. This points to a two-step process of irony processing. In terms of common ground, the two steps would be (a) agreement what we have in common, and then (b) accommodation of the new bits of information provided to enrich the already shared meaning. But for irony at least the first part has to be shared, as is described by Sperber and Wilson (1981) with their anchor in discourse that has to be present for irony. This is what I will discuss now, when comparing the concepts of presupposition, common ground, and irony.

2.6.2 Presupposition and irony

There are analyses, as Potts (2005) points out, that put all meanings that are not literal, he calls that “at-issue meaning” (Potts, 2005, pg 32), in a big category presupposition. He has various arguments against that claim but the crucial distinction between presupposition and implicature according to Potts (2005, pg 22) is as follows: The most important factor that divides meaning is deniability. He paraphrases this with the question: “Is p deniable in the context C ?” or “Is it possible that p is a potential, but not an actual, contribution to C ?” (Potts, 2005, pg 22). The Gricean conversational principle and its maxims guide the hearer to the most informative of the contexts the utterance U is compatible with. But if some other intended meaning of the speaker seems to be compatible as well, the hearer must be willing to regroup. So the speaker can deny one possible meaning by making a different conversational move (Potts, 2005, pg 22). Potts speaks of conversationally-triggered presuppositions that are close to the notion of presupposition that Stalnaker defined, as was described above, where it is allowed that some presuppositions are lexically triggered, too. They are according to his notion just like conversational implicatures context dependent and can be denied in some contexts. Conventional presuppositions, on the other hand, are determined by grammar and therefore always present when their trigger occurs (Potts, 2005, pg 23-24).

This is exactly the same as Heim (1990) writes: proponents of the view that presuppositions cannot be canceled understand them to work just like

conventional implicatures, that by definition cannot be canceled. They stand in contrast to conversational implicatures that depend on the context and are pragmatic and therefore can be canceled (Heim, 1990, pg 10,18).

Deleting presuppositions Frege stated that the reference of a sentence is its truth value and that the reference of a sentence is a function of the references of its parts, therefore if one part of a sentence has no truth value the whole sentence can't have one either. The same is the case for presupposition failure for one compound of the sentence, the whole sentence will then lack a truth value (Van der Sandt, 1992, pg 333,334). However, this prediction doesn't hold, and this problem is discussed under the name projection problem.

What I like to discuss here is the **referential value**, Van der Sandt (1992) points out to. What he takes from Frege is the different kinds of content, one evoked by the semantics and one by pragmatics. Presuppositions and implicatures equally enrich the meaning of a sentence but rather than being content that can be measured by truth value they are based on the propositional content (Van der Sandt, 1992, pg 336). Van der Sandt (1992, pg 341) proposes that presuppositions are anaphors with descriptive content and internal structure. He does not assume that presuppositions are canceled if they are not entailed by the context, rather they are flexible enough, since they have structure and content on their own, to repair discourse if no antecedent for the anaphor is present. So they accommodate and provide an antecedent themselves (Van der Sandt, 1992, pg 346). Both, propositional content and pragmatic information, need to be updated to the context (Van der Sandt, 1992, pg 336). Which is what is generally described with the notion common ground.

First, one has to outline that there are basically two points of views on presuppositions: The theories that understand them to work like **conversational implicatures**, and the theories that consider presuppositions as **admittance conditions**. The admittance conditions view is saying that the proposition presupposed must be one entailed in the common ground derived by the context. This view also assumes that if presuppositions add new information to the dialogue, as they can, the hearer readily adapts this new information to its common ground, a process called **accommodation** (Heim, 1990, pg 4,5).

The **projection problem** of presuppositions can be described with the following examples (Heim, 1990, pg 1):

- (37) a. John has a wife.
 b. John's wife is away.

- c. If John has a wife, his wife is away.
- (38) a. The king has a son.
- b. The king's son is bald.
- c. If the king has a son, the king's son is bald.

The examples in (37c) and (38c) are composed out of the sentences a. and b. they both presuppose that John exists and that he has a wife, or that there is a king and that the king has a son. But that John has a wife, or that the king has a son, is not presupposed by the whole sentence in c. So far we have understood Frege's principle of compositionality as working on every level and thus assumed that the meaning of a complex sentence is composed out of its component clauses and the way they are put together. That doesn't seem to work here. The presupposition entailed by its parts disappears when we put them together, it is not "projected up" (Heim, 1990, pg 24) to the more complex structure, therefore it is called projection problem. Rather, the *if* seems to delete the presupposition.

It is a much debated point whether presuppositions can be canceled. The proponents of the view in favor of the assumption that under certain conditions presuppositions are canceled, argue as follows: Since presuppositions entail the existence and uniqueness presupposition and entailments of the asserted content can not be canceled, there remains a contradiction if the presupposition is false. However, in a negated sentence, the presupposition of the affirmative remains unchanged but the content is negated and therefore does not entail the presupposition itself. What happens is that the presupposition clashes with the common ground and is canceled (Heim, 1990, pg 10).

The difference can be seen in the following negation (Heim, 1990, pg 9):

- (39) a. I talked to the king of France.
- b. I didn't talk to the king of France.

For negation it actually is assumed that presuppositions remain in a negated sentence. So both sentences presuppose that there is a king of France.

People that challenge that presuppositions can be canceled assume they work like conventional implicatures, that per definition cannot be canceled, in contrast to conversational implicatures. They explain the negated sentence above, that can be uttered without presupposing that there is a king of France, in that they argue that there are two different kinds of *not*. One where it negates the content while leaving the proposition intact, that is what I described in the last paragraph ("ordinary negation" (Heim, 1990, pg 10)). The other kind of *not* always derives a sentence without any presupposition

in the first place (called the "denial/metalinguistic negation" (Heim, 1990, pg 10)). If the sentence b. does not presuppose that there is a king of France they assume it is due to the metalinguistic-not. While the ordinary-not would yield to a sentence that presupposes that there is exactly one king of France.

At this point, I would like to point out that young people use the post-positive negation *not* without any inflexion to mark sarcasm, so that it is mostly used to express negative sentiments, explicitly.

(40) I would love to do that. - Not!

This works with any affirmative I think. The "not" has to be stressed therefore the exclamation mark. I have also heard this in German with the negation *nicht*. But it is also common to utter the phrase in German and use an English *not* afterwards, since this figure of speech apparently comes from the Anglo-American area. This inflected *not* as marker for irony (mostly sarcasm, since it derives from a positive meaning the negative meaning) does not only negate a certain part of the utterance, but negates the whole sentence and yield its opposite. Maybe this can help to understand how presupposition and sarcasm interact. Which kind of *not* would this use correspond to?

Maybe one can check this if we consider the following negated sentences.

- (41) a. The present king of France is not bald.
b. The present king of France is bald. - Not!

We saw above that the presupposition in a. holds even for the negated sentence. If for the sentence in b. the presupposition remains intact the *not* could be described as ordinary-not, while if it does not presuppose that there is a king of France it is a metalinguistic-not. I would say the former is the case: the presupposition remains for the ironic use. According to my own intuition a. and b. still presuppose that there is a king of France, he is just not bald. This would make the colloquial sarcastic use of the post-positive *not* just an ordinary negation and not a metalinguistic one. But more importantly this makes sense since presuppositions are closely linked to the notion of common ground and there is no doubt that shared background knowledge is a central point for irony: To enrich meaning one needs to know what is already shared believe, as we saw already for jokes.

I wrote above in (2.5.1) that it is difficult to argue with a concept that is that much topic of debate like the idea of a common ground. However, there cannot be a doubt that it plays an important role in describing how an ironic meaning arises.

I touched upon the issue already in saying that we can presuppose something that is not in the common ground previous to the dialogue, because we expect our hearer to accommodate. This does not work with irony. An ironic issue has to be present in the common ground prior to the ironic utterance. So for presupposition I can expect my opposite to update the presupposed information that sets the common ground plus the new information uttered at the same time. For irony, on the other hand, first, the common ground has to be secured and updated with all that is presupposed and then, second, the new ironized issue can be added. That argumentation is also in favor of a two-step process of irony processing.

2.6.3 Irony as enrichment of meaning

Before I will turn to psycho- and neurolinguistic studies that can provide evidence for the two-stage view argued for above, I would like to sum up the claims that I have made in the discussion so far and elaborate on one possible interpretation of it.

Bringing together what I argued about irony and implicatures and what we saw now in terms of presuppositions and common ground, what is the mechanism then that makes irony function as enrichment of meaning that makes the previous adaptation of common ground between the interlocutors necessary? And how does this together with what we perceive as funny?

One part of the answer might be that there are different kinds of enrichment.

Chierchia (2006, pg 2) writes about Polarity Sensitive Items that can be divided in two types: negative polarity items (NPIs) and free choice items (FCIs). There are different distributions for words and their function in different languages. *every* for example functions as free choice item while *any* can be both FCI and NPI. There has been described a communicative function to describe the syntax and semantics of polarity sensitive items, that is domain widening: “It is well known that as we communicate, we select domains of discourse as our subject matter. Non referential DPs like *every student*, *a student*, etc. are used with such domains in mind” (Chierchia, 2006, pg 3, italics in the original). That means we assume someone is speaking of *every student* in a relevant domain in this certain context and not about every possible imaginable student there is. So we restrict the entities we are talking about as interlocutors to a domain that seems relevant to us and the topic of discussion. NPIs now, for example *any* compared to *some*, “contain an instruction to consider domains of individuals than what one would otherwise do.” That means they contain an implicit instruction to widen the domain considered for the current discussion. What domain

widening does in negative contexts is that it leads to a stronger interpretation because it excludes more alternatives. In a positive context, on the other hand, “widening the domain of an existential leads to a statement which is *weaker* (i.e. less informative) than what we would obtain with a plain indefinite“ (Chierchia, 2006, pg 3, italics in the original). Therefore domain widening explains why in fact NPIs make more sense in a negative context.

Chierchia (2006, pg 3) writes as an example:

(42) *There is any student in the building.

(43) There isn’t any student in the building.

To sum up, “within the scope of negation [...] consideration of a broader domain leads to a stronger (and hence more informative) statement” (Chierchia, 2006, pg 3). The entailment relation of affirmative sentences is upward entailment, it “licenses inferences from subsets to supersets” (Chierchia, 2006, pg 3). For example: My friend likes to smoke Gauloises, entails that she likes to smoke cigarettes. Which brings us to downward entailment, which is “the capacity to license inferences from supersets to subsets” (Chierchia, 2006, pg 3). For example: If I smoke, I feel sick. Which entails: If I smoke Gauloises I feel sick. The communicative function of domain widening of NPIs is explained by the logical principle of downward entailment. That serves the communicative practice to “maximize information content with parsimonious use of resources” (Chierchia, 2006, pg 4).

This is an example of how pragmatic principles work together with grammatical mechanisms, or use them for their purpose. So how does this mechanism work? Quantifier can form a scale, in that the use of *all* entails *many* which entails *some*. So by saying *many* I decided not to use the stronger one and not the weaker one. The listener can conclude now, by using Gricean reasoning, that I want to communicate something stronger than *some* but not the stronger *all* which is exactly the scalar implicature that comes with quantifiers, for example *some* can be pragmatically interpreted, that is enriched, as *some, but not all*. As Chierchia (2006, pg 10) formulates it: the interpretation of the hearer is “typically guided by the awareness that one could have made weaker or stronger assertions. In particular, scalar items seem to automatically activate the alternatives constituted by their scalar mates.” What is interesting here, is the “leap of faith” that Chierchia (2006, pg 10) assumes about the state of information that is necessary besides Grice’s maxims and logic. These kind of entailment structures that constitute the scale are also a form of “likeliness” (Chierchia, 2006, pg 10). Therefore one can speak of enrichment of meaning, because by mentioning a quantifier in a scale the alternatives come to mind automatically and therefore their meaning adds to the communicated information, too.

But strengthening the meaning via downward entailment is not the only way in which implicatures can change the informational state of an polarity item. Extreme uncertainty can also be derived, for example with the German *irgendein* (Chierchia, 2006, pg 29). This relates to the two types of FCIs: an existential meaning, where *exactly one* is meant by *any* and an universal meaning where *any* means *all* (Chierchia, 2006, pg 6).

For example, like in Chierchia (2006, pg 30):

(44) I saw any student that wanted to see me.

Here the universal meaning is the relevant one: the sentence must be true for every possible student that wanted to see me. This reading puts the assertion and the implicature together to form a meaning (Chierchia, 2006, pg 30). Alternatives lead, if present and relevant in a certain context, to enrichment (Chierchia, 2006, pg 14). So when do alternatives in a certain context lead to strengthening and when to uncertainty? Chierchia (2006, pg 30) concludes that there are two kinds of implicatures, the NPI implicature as a case of even-implicature with an universal reading that leads to uncertainty and the FCI implicature as antiexhaustiveness that has an existential reading and leads to strengthening. Exhaustification is an operation where *only* compares to relevant alternatives, so the existential reading of *any* for example is derived as *exactly one*. Antiexhaustiveness means the universal reading is derived, in the case of *any* a meaning like *every*. To sum up so far, enrichment takes place and with scalar implicatures it leads to strengthening, but not in all cases, there are also cases of enrichment that don't lead to strengthening (Chierchia, 2006, pg 17).

"Negation reverses entailment. Hence the assertion becomes the strongest of all alternatives. Every single alternative is entailed by the assertion" (Chierchia, 2013, pg 174). As we have seen already: "Scalar implicatures associated with not-end-of-scale items in negative contexts produce positive implicatures for any subdomain of the original domain" (Chierchia, 2013, pg 175). And therefore: "End-of-scale existential operators yield weakly informative statements in positive contexts and strong ones in negative contexts. This, in some obvious sense makes them 'more useful' in the latter." (Chierchia, 2013, pg 183).

He concludes that FCIs have to have three things (i) existentiality, (ii) an antiexhaustiveness implicature, and (iii) a scalar (uniqueness) implicature (Chierchia, 2006, pg 37). Furthermore, he introduces a "freezing operator" (Chierchia, 2006, pg 42) that comes in two forms, a strong, presuppositional one and a weak, non presuppositional one. So strengthening goes along with the strong presuppositional operator that applies to something positive. And

therefore FCIs in a positive context acquire an existential reading (Chierchia, 2006, pg 43).

To conclude, domain widening requires the activation of relevant alternatives, those alternatives trigger implicatures. This is the point Chierchia (2006, pg 43) takes from Grice: “a conversational move is judged against a background of other a priori conceivable moves. Selecting a move over another can be very telling.” This is how interlocutors enrich communication in a very efficient way, by using the underlying principles of communication in a recursive, compositional manner, constantly and systematically throughout the computation of meaning (Chierchia, 2006, pg 44).

If the alternatives now form a scale (i.e. a linearly ordered set), then choosing an element will naturally indicate that all alternatives which are not entailed are not deemed to hold. [...] If the alternatives do not form a perfect scale [...] we ought to choose the one which enables us to make the strongest (and hence least likely) statement; accordingly, the hearer, making the usual leap of faith, will conclude that that is indeed what is intended and an “even” implicature naturally comes about. [...] If, on the other hand, we are excluding no alternatives of any size, down to the smallest possibility, then it sounds like we are really uncertain; we ought to choose, therefore, the assertion that commits us the least, the one that enables us to rule out fewer possibilities. From this, the hearer will jump to the conclusion that the speaker is trying to rule in most possibilities (and hence the existential statement being made is likely to hold of *every* alternative) (Chierchia, 2006, pg 44).

For Grice every sentence is uttered “against a background of possible alternative statements” (Chierchia, 2013, pg 102), which is something especially interesting for irony. Where the uttered sentence clashes with the present background and the hypothesis is that the hearer is forced to derive a different background against which the utterance actually makes sense. This absurd or at least contradictory background in light of the real context is what is funny. The logic behind Grice’s approach is that what is said and what wasn’t said contributes to meaning and by reasoning about that the hearer “gathers insight on the communicative intention of a speaker” (Chierchia, 2013, pg 102) this exactly will be the issue of the next chapter about studies in (3.3).

So in light of Chierchias framework for implicatures as well as for irony this following principle holds: “[T]he principle that when alternatives are

active they must be factored into meaning” (Chierchia, 2013, pg 185). To exhaustify means for Chierchia run the Gricean reasoning (Chierchia, 2013, pg 106), as follows (Chierchia, 2013, pg 101):

- (45) a. Joe or Bill will show up
- b. Joe and Bill will show up
 - 1. The speaker said (45a) and not (45b), which, presumably, would have been also relevant. [relevance]
 - 2. (45b) entails (45a), and hence (45b) is more informative than (45a).
 - 3. If the speaker believed that (45b) holds, she would have said so. [quantity]
 - 4. It is not the case that (45b) holds.
 - 5. The speaker has an opinion as to whether (45b) holds. [opinionatedness] Therefore:
 - 6. The speaker takes (45b) to be false.
 - 7. The speaker is conveying that John or Bill and not both will show up.

By a line of reasoning like this in the end the classical implicature of *or* implying *but not both* is derived. One could also call this, as Chierchia (2013, pg 106) does, maximizing strength. Since *and* is logically stronger than *or* uttering *or*, which is entailed by *and*, is more informative.

Chierchia’s own approach of implicatures as exhaustification is constructed in parallel to Grice’s approach: “the Gricean reasoning [...] can be viewed as inducing an exhaustification of the assertion with respect to the relevant alternatives” (Chierchia, 2013, pg 102). The problem with Gricean reasoning has always been that as intuitive as it is there is no formal way to model it (Chierchia, 2013, pg 109). So what Chierchia (2013) proposes is that exhaustification is defined as applying a covert *only* (Chierchia, 2013, pg 100). This means that the hearer assumes that the exact assertion uttered is the **only** one that counts. But besides *only*-exhaustification (O) there is another option how what has been said can activate alternatives, namely *even*-exhaustification (E). Chierchia’s example is *everybody* (Chierchia, 2013, pg 109):

- (46) a. Mary went to the party, greeted everybody, hugged Paul and Sue and left [*only*-exhaustification]
- b. Really everybody came to my party. Imagine that MY EX came. [*even*-exhaustification]

From a Neo-Gricean perspective, these operators locate what was said relative to relevant alternatives on a likelihood scale. And “we simply use whichever of these operators works best to enrich the assertion, given the alternatives that are made salient through focus/context” (Chierchia, 2013, pg 110). That counts for scales ordered by entailment. But there are also pragmatic scales, which would be the ones interesting for irony. The attempt is to “generalize scalar reasoning to pragmatically supplied partial orders” (Chierchia, 2013, pg 111). Unfortunately Chierchia (2013) doesn’t have an answer, yet:

“The point is simply that some entailment patterns are stably rooted into pretty universal logical formatives (like the quantifiers and the numerals); others may be linked to classes of predicates subject to a greater degree of cultural and contextual variation across languages. This is a far cry from a wholly worked out theory of pragmatic scales. But it gives some ground for optimism about the role exhaustification might play in the way we reason about them.” (Chierchia, 2013, pg 112)

“The general idea is that NPIs are low elements on a scale of operators/quantifiers. They obligatorily activate their scalemates and, for appropriate items, domain alternatives. This triggers implicatures responsible for their distribution. Implicatures arise as they do, in general, via exhaustification through O or E” (Chierchia, 2013, pg 182-183). These are two different exhaustifiers or modes of exhaustification. Chierchia (2013) distinguishes between non-emphatic NPIs like *only*, this would be the exhaustifier O. And emphatic NPIs like *even*, which he calls even-exhaustification, or E. “*Even* (E-)exhaustification is more strongly emphatic than *only* exhaustification” (Chierchia, 2013, pg 148). What means emphatic? For *even* as E-exhaustification a set of linearly ordered alternatives is obligatorily activated. While for *only* as O-exhaustification alternatives form a “lattice structure” (Chierchia, 2013, pg 165). For *even* alternatives are ranked along a scale that makes a domain most likely. For *only* the probability ranking remains (Chierchia, 2013, pg 165). So they differ in terms of the probability measure which determines how domains are ranked in terms of their likeliness to find relevant alternatives in a certain domain (Chierchia, 2013, pg 158).

Where does this leave us? Chierchia (2013, pg 114) does not want his approach understood as a real alternative to Grice’s. His operator O has nothing to say about relevance implicatures. Also, the areas of irony, metaphor or which reading might be the one relevant in a certain context, “remain untouched by covert exhaustification” (Chierchia, 2013, pg 114). The idea still is intriguing: Chierchia (2013, pg 104) speaks of “relevance implicatures”

that are highly context-dependent where “one seeks to accommodate what looks like a blatantly irrelevant move” in contrast to scalar implicatures, where the scales are lexically determined. Irony would clearly belong to the former type, where a possible scale can only be pragmatically determined.

Having said that one can carefully phrase a few hypotheses:

1. Scales are the formalization of pragmatic background used for implicatures and relations of entailment. But there are different entailment relations.

2. First, there are entailment relations that function logically as we saw with scalar implicatures for example with numerals (5 entails 2) that are pure semantic scales, they are semantically driven and follow pure logic. Second, there are pragmatic scales that construct a likeliness scale as background, that is, pragmatic information in terms of likeliness, created by knowledge, ordered by disjunctive conjunction (Grice (1967, pg 58) said irony is created by “disjunction of such specific explanations”). In other words, linking different parts of the scale, which would be a likeliness scale with entailment, similar to the one of *even*. At the logical level the scale has to be valid, which is to say, the entailment relationships have to be valid.

3. Irony is playing with the likeliness of entailment, exploits the likeliness created by an absurd context or background. The literal meaning is true relative to the new background that to be changed.

4. Therefore, the hearer is forced by the speaker to create an absurd context against which the utterance makes sense. The creation of this background is irony and amuses us. This works like soft presupposition triggers, that is, accommodation (not the kind we talked about for presupposition), I have to adapt the context not the utterance to force the utterance to make sense. That is why the common ground (unlike for presupposition) has to be settled beforehand for irony. Because irony forces the hearer to accommodate a different absurd background from it. One can imagine this like the figure-ground reversal as logical mechanism for jokes (with the light bulb and the poles that turn the table), the ironic utterance remains steady but the background, it has to be incorporated in, needs to be changed accordingly.

5. The negation or opposite of the literal utterance merely follows trivially from that and it is not forcibly part of the irony any more. The reasoning after the accommodation of context is what denies the literal utterance as a third step of the grounding process.

6. If you don't get the irony, you do interpret it literally but in a context that is absurd, but you are not denying the utterance in the logical sense. When constructing the appropriately absurd context to make sense of the utterance it follows from that that the literal utterance is denied, by the hearer, not the speaker. That is important to explain why irony can be like

a conversational offer, where it is in the hands of the listener to construct the fitting background, or context, or likeliness scale, however you might want to call it.

Coming back to the example above: *She is smart and beautiful.* -*Yes, she is smart.* One could now assume that the implicature in the answer, which can also be used ironically forces the hearer to construct a background, where she is **only** smart, which is exhaustification, and excluding the alternative of her to be smart and beautiful. The same was the case for our example with Gricean reasoning with *Joe or Bill showing up* which through constructing the implicature excludes the alternative that both will show up, which would be the stronger claim since *and* is logically stronger than *or*. Just like that *smart* is logically stronger than *smart and strong*.

Problems The framework of Chierchia (2013) might provide some insight for some instances of irony, but there remain some problems with this idea. First of all, the problem is that there does not exist a theory of pragmatic scales, which we would need to talk about how exhaustification works with irony. Therefore many questions remain: If it is completely pragmatic where does the scale come from? Which axis is the scale to put on? Is the strength in terms of deviation of the literal meaning also ordered on a scale? If there are too many variables how can the interlocutors construct a common scale? There might be kinds of irony that work with different logical mechanisms, rather than scalar ones, for example, we saw that some jokes use figure ground reversals. How should the alternatives be ordered then? Following the reasoning of Chierchia (2006, pg 44), if the alternatives do not form a linearly ordered set, the implicature triggered cannot be one that strengthens meaning but leads to uncertainty in the case of irony. So irony would be one of the cases where enrichment doesn't lead to strengthening as Chierchia (2006, pg 17) mentions. Also why does irony, as sarcasm, function better in a positive context, just like FCIs do? Sarcasm is saying something positive meaning something negative, negation creates a downward entailing context, which normally is for NPIs that lead to strengthening. Maybe sarcasm leads to strengthening of meaning while kind irony leads to weakening?

In the end of this section, another question is, which points to the following chapter about psycholinguistic evidence for theoretic approaches in chapter (3): What if, while scalar implicatures are a strengthening device irony is a weakening device? Chierchia (2013, pg 24-25) writes: "Weakening an assertion (through the addition of an implicature) is not impossible. But we do so only when the (pragmatically unenriched) sentence would overtly clash with the context at hand". This is a very good summary of the hy-

pothesis discussed above. Irony as a pragmatic device to weaken meaning, this is called the “tinge hypothesis”, which Filik et al. (2015) in their study discuss, and we will come to this later in (3.4).

Chapter 3

Psycho- and neurolinguistic studies

In this section, I want to present some hopefully illuminating psycho- and neurolinguistic studies on what was discussed so far from a theoretical point of view. It is said for children, autistic people and people suffering from right-brain damage after strokes that they have problems with the comprehension of humor and irony (most often, sarcasm is tested) especially. That is why these three populations take an important place in the research on the comprehension of humor and therefore also in this chapter.

I will structure the discussion as follows:

First, I would like to show evidence for a two-stage model of irony processing. Studies suggest, to fully appreciate the mismatch between literal and non-literal meaning both interpretations need to be accessible to compute the difference between them. What essentially makes us laugh is the contradiction with the real world, which made a reinterpretation necessary. Second, as I used a part of the proceeding chapter to show similarities and distinctions among presuppositions, implicatures, common ground, and irony, an issue deeply linked to pragmatic meaning is the idea of a theory of mind, which allows us to attribute mental states to others. This leads to suggestions that the more general part of communication related with this is making our intentions clear. Finally, many studies claim that the comprehension of irony is related to the more general skill of communicating intentions. Here, I will mostly focus on studies with people with autistic spectrum disorders (ASDs).

Irony is deeply embedded in a certain context, therefore one has to mention that psycholinguistic studies can do justice to that only to a limited amount. For example, for written irony the type of text can be crucial: We would accept an ironic meaning such like sarcasm in a commentary with more ease than in an report that needs to be objective and neutral. Or for verbal

irony, a lot of the clues we use in detecting and decoding it come from our personal relationship with the speaker. Asking participants to judge sentences in a psycho- or neurolinguistic design, in most cases, cannot depict these kinds of contexts. But I will present some studies that try to incorporate for example a relationship between the interlocutors in their study-design, or record irony in natural speech.

3.1 Two-step interpretation

First, as this turned out to be a central claim of the discussion of theories concerned with the comprehension of irony, I would like to present some evidence for a two-step account of irony comprehension.

What follows are two studies that have been mentioned already in the discussion of theoretical concepts on irony. Here, I would like to focus on another aspect of theirs, namely, the two-step process that is assumed to take place for irony comprehension.

Filik and Moxey 2010 findings support the graded salience hypothesis and a two-step process of irony comprehension where both interpretations remain active. Their eye-tracking study had the aim to investigate in how many stages it takes to comprehend irony: "If a mismatch with the context indicates that the literal interpretation is inappropriate, it is necessary to cancel the surface-literal interpretation, and compute the non-literal interpretation by assuming the opposite of the literal interpretation" (Filik and Moxey, 2010, pg 421). To recapitulate, this assumption is in line with the standard pragmatic model (following Grice, 1967), where different processing costs for literal vs. non-literal language are assumed. The direct-access view, on the other hand, presumes that figurative and literal language are processed the same way, in that the context interacts with language processing very early on and allows to decide straight away which interpretation fits and needs to be computed. In between the two stands the graded salience hypothesis, where the most salient interpretation, might it be the figurative or the literal, is accessed first. Salient means encoded in the mental lexicon and prominent in their frequency, conventionality, familiarity, or prototypicality. The standard pragmatic view assumes that for irony the overwritten literal interpretation is canceled. The graded salience view comes with the assumption that both interpretations remain accessible to compute the difference between them, since that is what irony does, it "draws attention to a failed expectation" (Filik and Moxey, 2010, pg 422). The study design to decide between those claims looked as follows: To see which interpretations

remain active they used pronominal reference, since positive and negative quantifiers focus on different sets of discourse entities. Irony should inverse the focused pattern of reference, since it inverts meaning. In investigating which references are activated, one can check if an ironic meaning has been computed or a literal one and one can check which references, as sets of discourse entities, remain active. The following sentence is an example from Filik and Moxey (2010, pg 423):

- (47) a. *I see many people have come to your party. They heard about the free drinks.*
 b. *I see many people have come to your party. They couldn't get a babysitter.*

If a. is uttered in a context where the party is crowded, it would be meant literally and the *they* would refer to the set of all the people being there. The sentence in b. uttered at a party where there is only few people there, would be ironic and the *they* would refer to the set of all the people not there.

They found that ironic sentences indeed required more processing effort, measured in longer reading times. Which is in line with the standard pragmatic view and against the direct access view. The processing of pronominal reference suggests that readers had access to the discourse entities evoked by the literal and the ironic interpretations of the quantifier. This is not in line with the standard pragmatic view in so far, as it suggests that one interpretation is replaced by another and no longer accessible. Both findings support the graded salience hypothesis and the functional principle that the retention hypothesis describes, namely, that both interpretations remain active so that the difference between the two can be computed to draw attention to the failed expectation (Filik and Moxey, 2010, pg 428). They conclude, both interpretations are required as parts of the ironic meaning (Filik and Moxey, 2010, pg 430).

Giora et al. 2000 investigated in the graded salience hypothesis, too. They conducted a study with brain damaged patients and their understanding of irony. They also were interested to find psycholinguistic evidence against the view taken by relevance theoreticians (for example, Sperber and Wilson, 2001) that if the context is rich enough figurative language such as irony is processed with as much ease as literal language. Against that, they hold the view of traditional theories following Grice that assume irony requires a special processing that involves more cognitive effort.

On the traditional view, comprehension obligatorily involves activating the literal meaning first. Nonliteral interpretation is de-

rived only when the literal meaning has failed to meet contextual fit. Understanding nonliteral language must, therefore, involve more complex inferential processes, regardless of context (Giora et al., 2000, pg 64).

To test the graded salience hypothesis that predicts a sequential process for "nonconventionalized, nonsalient sarcasm" (Giora et al., 2000, pg 65) and considering that the right-hemisphere is especially involved in linguistic reinterpretation, they conducted a study with right-brain-damaged (RBD) patients and their understanding of sarcasm. The hypothesis was, in line with the graded salience hypothesis, that there is a special involvement of the right hemisphere in the comprehension of nonsalient sarcasm. This was compared to the comprehension of salient metaphor and left-brain-damaged (LBD) patients. In distinguishing sarcasm from lie, the context was such that a person was saying something sarcastically to another person that had the same information than the speaker, so there was no point in lying to that person because he could tell that it would be a lie (Giora et al., 2000, pg 66). All the impaired participants performed worse in rating the utterance as sarcasm compared to the unimpaired control group, however, there was no significant difference between patients with left and right brain-damage (RBD patients performed slightly worse than LBD patients, though) (Giora et al., 2000, pg 67). But when the results were cleared from the effect of aphasia, that means the effects of general language problems were removed, RBD patients performed significantly worse than LBD patients (Giora et al., 2000, pg 68). To make sure that this was not due to the fact that LBD patients could better interpret the prosodic clues of the sarcastic utterance, their performance on the prosody subtest was included in the analysis, showing that there was no significant difference in the prosody subtest between LBD and RBD patients (Giora et al., 2000, pg 69). Comparing those participants that did bad on the prosody subtest with those that did well revealed no influence of prosody in the understanding of sarcasm (Giora et al., 2000, pg 70).

Interestingly, among the patients that understood the passages but still chose the wrong answer most common was the "lie"-choice, while for those that did not understand the passage, indicated by questions to text comprehension in general, the most popular incorrect answer was "literal truth" (Giora et al., 2000, pg 71). The authors conclude that those who understood the passage must have a theory of mind, since they chose "lie" incorrectly as answer, which means they knew that the speaker had some intention in mind, otherwise they would also, just like the participants that did not understand the passage, have chosen the "literal truth" answer (Giora et al.,

2000, pg 73). Summing up, the data suggests that both, right and left hemisphere, are involved in sarcasm comprehension, but the involvement of the right hemisphere seems greater (Giora et al., 2000, pg 72). Contrary to this, the comprehension of conventional metaphor was significantly impaired for LBD patients compared to healthy controls and RBD patients (Giora et al., 2000, pg 76).

The graded salience hypothesis predicts nonsalient sarcasm should involve the right-hemisphere and the understanding of meaning, that is coded in the mental lexicon such as salient conventional metaphors, should involve the left-hemisphere. So, the authors see their hypothesis confirmed in that processing of language requires different strategies depending on the salience of the meaning that is to interpret.

Summary Both studies provide evidence for more processing effort for ironic language compared to literal language. To detect the mismatch between what is meant and what is said, the literal and non-literal interpretation have to be required to detect the failed expectation, just as to detect opposition one needs both of the opposing views. The results are more in line with the standard pragmatic view and the graded salience hypothesis than the relevance theoretic account. And what became clear: investigating in the processing effort of irony can hardly be separated from having a concept like theory of mind. Which will be the issue in the next section.

3.2 Theory of mind

What became clear now, is that one cannot discuss the psycholinguistic view on irony without talking about theory of mind. A theory of mind is defined according to Baron-Cohen et al. (1985, pg 38) as ability

to conceive of mental states: that is, knowing that other people know, want, feel, or believe things [...]. A theory of mind is impossible without the capacity to form ‘second-order representations’.

Baron-Cohen et al. (1985, pg 43) proposed that children with an impairment within the autism spectrum disorder (ASD) lack second-order representations and therefore do not have a theory of mind and also lack pretend play. They conclude: “The ability to make inferences about what other people *believe* to be the case in a given situation allows one to predict what they will do” (Baron-Cohen et al., 1985, pg 39, italics in the original). Since then, the question whether one can explain pragmatic deficits with a lack of theory of

mind has been the interest of research. They provided evidence that children within the autistic spectrum disorder lack theory of mind. Furthermore, they categorize this as a cognitive mechanism independent of IQ. Therefore, they have been the first to describe the deficit of ASD children with a more general cognitive failure (Baron-Cohen et al., 1985, pg 44). This cognitive mechanism and its role for the computation of the meaning of irony will be the issue of this section. I will come back to the special impairment of ASD in the section that follows in (3.3.1), bringing together theory of mind and patients within the autistic spectrum disorder.

Theory of mind is not always the same, it comes in different forms: first- and second-order. Second-order theory of mind is the more complex more embedded version of first-order theory of mind. If theory of mind first-order would be to reason about another person's state of mind, second-order is reasoning about another person's state of mind that is reasoning about something. A first-order thought: I think about what you think. A second-order thought: I imagine what you think that I think. I will now summarize a paper by Reimer (2013) that I think is exemplary for the discussion about different theories that make different predictions about what kind of theory of mind is needed to interpret non-literal speech.

Reimer 2013 Reimer (2013) argues that empirical data, indeed, supports Grice's view of irony as figure of speech and therefore the standard pragmatic view. She presents counterarguments against studies, such as the one from Gibbs (2000, pg 2), that argue against Grice's view on irony and are in favor of relevance theoretic accounts. She reviews Grice's idea of irony comprehension as the hearer recognizes that the speaker flouts the maxim of quality and understands the conversational implicature that the expressed proposition is "contradictory" to the uttered proposition: "This must be some obviously related proposition; the most obviously related proposition is the contradictory of the one he purports to be putting forward" (Grice, 1967, pg 53). This is a quote that I cited already when talking about the theories in (2.1). Reimer now concludes, if irony conversationally implicates the contradiction of what was said, the initial interpretation has to be replaced by its contradiction. A process she describes as involving **second-order theory of mind**. Which might be nothing else than what Sperber and Wilson called a "second-degree interpretation" (Sperber and Wilson, 1995, pg 238), as we have seen in the chapter on the echoic mention theory in (2.2). Reimer defines theory of mind as "the ability to attribute mental states to others in order to explain or predict their behaviour" (Reimer, 2013, pg 6). With first-order theory of mind she means attributing mental states about the world

to someone else, while second-order theory of mind means attributing mental states about the mental states of others to your opposite. Researchers (for example, Happé, 1993) claim that the relevance account predicts the comprehension of metaphor to involve only first-order theory of mind while for irony it predicts that its comprehension requires second-order theory of mind. Grice's account on metaphor and irony, as they say, predicts for both to require second-order theory of mind. Therefore only the former account can explain the differences found in psycholinguistic research between understanding irony and metaphor. Now Reimer (2013) is defending Grice's view on irony and metaphor in saying that the experimental evidence on metaphor (conducted with people that had only first-order theory of mind and people that had first- and second-order theory of mind) that is supposed to be not in line with the predictions of the Gricean approach, is not credible (Reimer, 2013, pg 9). She argues, the metaphors used in the tests, whether metaphor requires first-, or second-order theory of mind, are dead metaphors. That means they are already part of the conventional interpretation of a phrase and make the phrase lexically ambiguous. In these cases, it is no wonder that empirical evidence (that I will present later in (3.3.1) suggests no second-order theory of mind is needed to interpret these metaphors. Other tests (see Happé, 1993), were rather a test of encyclopedic knowledge that is needed to match a metaphor to a description and not to understand the metaphorical meaning of an utterance (Reimer, 2013, pg 11-13). In the end, she argues that metaphor and irony require a second-order theory of mind to be understood and therefore the account of Grice would make the right predictions and not the relevance theory that predicts different theory of mind-requirements for the two instances of non-literal speech. Again, what Grice claimed was that irony such as metaphor involve conversational implicatures that are defined as depending highly on the context (Reimer, 2013, pg 14). In her opinion, the used examples of metaphor lack a particularized conversational implicature or do not require the participants to interpret it and it is this implicature exactly that needs second-order theory of mind to be interpreted (Reimer, 2013, pg 16). Actually Reimer (2013) goes so far as to say that for Grice every real understanding of language requires second-order theory of mind. Anything else (also the instances investigated in the studies) would be merely decoding language, since to understand all natural speech we need to interpret what the speaker meant (Reimer, 2013, pg 17, 18).

By looking closely on irony in the Gricean sense, Reimer suggests that interpreting irony even requires second-order theory of mind employed twice: Once, to rule out the literal interpretation and the second time, to interpret the intended meaning (Reimer, 2013, pg 22).

Nakassis and Snedeker 2002 Nakassis and Snedeker (2002) investigated in irony comprehension among healthy children to understand how humor processing evolves. What they were most concerned with is the total absence of a consensus about the time when children start to understand jocular speech. They explain this disagreement among researchers with a lack of sharp divisions in the question itself: What does it mean to understand irony or its special case sarcasm? Do children understand sarcastic speech if they can label utterances with "nice" or "mean"? Or do they have to be able to construct the intended meaning in contrast to the literal meaning? The latter, they claim, appears at the age of around six. A third possibility to test irony comprehension is to ask children to give "second order believe judgments" (Nakassis and Snedeker, 2002, pg 429), this is the ability that research finds to emerge the latest. They stress, however, that this ability concerns metalinguistic understanding and has nothing to do with how irony comprehension naturally takes place. This can be understood like the implicit knowledge about language and explicitly explaining it, which are two very distinct things, where for the former the latter is not necessary. They hold the view as echoic mention theory takes it that irony expresses some kind of attitude towards the issue, and they follow this approach in that it is also assumed that some instances of irony can be constructed as quickly and easily as literal language. Most developmental research on ironic speech has focused on sarcasm, since it is very frequent and clear cut. They argue that ironic compliments are more sophisticated and typically aimed at adults. Since they adopt the echoic theory irony most easily echoes social norms that are normally positive and that is why sarcasm is more common. Ironic compliments require a clear antecedent in discourse and are therefore more sophisticated (Nakassis and Snedeker, 2002, pg 431). What they are interested to find out is whether it is easier for children to detect sarcasm, that is ironic compliments (that are meant as insults), than ironic insults (that are meant as praise). Furthermore, they investigated in the role of intonation for children. It was examined how these two variables changed at different ages of the children (Nakassis and Snedeker, 2002, pg 430, 431, 433).

They tested children at the age of 5 years and 10 months up to 7 years and compared them to adults. What they found was that for adults and children irony was more often understood if the context provided clear motivation for the ironic interpretation. If the context did not provide very clear clues, adults were better than children in detecting irony. There was no difference between comprehending sarcasm or kind irony for children or adults (Nakassis and Snedeker, 2002, pg 435). They conclude that children at the age of six seem to be able to interpret some irony, that is sarcastic

utterances and ironic compliments. Children and adults seem to be using prosody as additional clue to the same amount. Children were more likely to revise their interpretation if the context required it, if they did not understand the irony they hesitated to answer, indicating that in these cases they knew something was going on. According to echoic theory, sarcasm should be understood more easily than kind irony. This was not the case here (Nakassis and Snedeker, 2002, pg 437). By contrast, children seemed not to distinguish between ironic insults and ironic compliments (sarcasm) since their irony comprehension is not yet based on metalinguistic knowledge. "Clearly, this pattern of performance suggests that children who are just beginning to comprehend the speaker's meaning have already formed a strategy for interpretation that works equally well for ironic insults and ironic compliments" (Nakassis and Snedeker, 2002, pg 438). They conclude that comprehension of irony involves a general understanding of this kind of language use, since it cannot be explained by frequency or convention that the children might have already acquired. Because then sarcasm, as the more frequent form of irony, should be easier to interpret than kind irony. Intonation showed to be not the most important clue, children and adults used it for irony detection but not to the same amount as they relied on the context. Even children at the age of six were able to interpret prosody in relation to a special meaning. However, they rule out that there is such a thing as a distinct ironic tone of voice.

This study, thus, does not provide evidence for the echoic mention theory that suggests to treat sarcasm and irony as two distinct concepts. Metalinguistic understanding of sarcasm and irony might be present in adults interpretation of irony but it is not in children and still, they seem to have an idea about figurative language when the literal interpretation mismatches the context. The following studies investigated in the processing of irony with patients suffering from right-hemisphere brain damage or Parkinson.

Winner et al. 1998 Winner et al. (1998) was the first study to focus on irony comprehension among adults suffering from right-hemisphere brain damage after stroke. They understand the ability to make **second-order mental state attributions** as one component of the theory of mind and investigated whether patients with right-hemisphere brain damage are impaired in distinguishing intentionally false ironic jokes and lies and whether such a difficulty is related to problems in attributing second-order mental states to others (Winner et al., 1998, pg 91). The second-order beliefs were either true (for irony, the speaker does know the hearer knows) or false (for lies, the speaker doesn't know the hearer knows) (Winner et al., 1998, pg 92). After

the participants were given a written story they had to answer different questions to check their understanding. First-order believe questions asked for judgments of a person's believe about the world, while second-order believe questions asked for a person's believe about another person's believes. There was no difference between the right-hemisphere brain damaged patients and the healthy control group for the first-order believe question (Winner et al., 1998, pg 96). Asked about the second-order beliefs the patients with brain damage made more errors than the control group on both the lie- and the joke-stories (Winner et al., 1998, pg 97). Asked about the speaker's expectation in the stories the patients did not make more mistakes than the healthy controls. In the final interpretation question, where participants were asked to judge whether the utterance was a lie or a joke, the patients with right-hemisphere brain damage made significantly more errors than the healthy control subjects. The relationship between second-order belief and final interpretation was similar among patients and healthy controls: "high error scores on second-order belief predicted high error scores on interpretation" (Winner et al., 1998, pg 98). Since the patients were more impaired for the first measurement they also were for the second. The participants also had to answer second-order expectation questions of the sort: does the speaker expect to be believed. The errors in this task were also strongly correlated with the final interpretation whether an utterance was a lie or irony. They conclude from their findings that their thesis was confirmed: the ability to attribute second-order mental states to others is related to the ability to distinguish lies from jokes depending on what I assume about the knowledge of the speaker about the mental state of the hearer (Winner et al., 1998, pg 99). The patients did not perform worse than controls on the second-order expectation question, especially for the joke-stories. This might be due to the fact that the person making a joke, correctly attributes an appropriate second-order expectation, while the liar incorrectly believes to be believed, which is an inappropriate second-order expectation. The latter might be more difficult to attribute. Against this interpretation that lies are more difficult to interpret than jokes, however, point the results by Monetta et al. (2009). This is the study we will turn next to, where they found that impaired participants chose wrongly the lie choice over a joke ending. In general, even though they conclude that difficulties in attributing mental states to others found in the study are clearly due to right-hemisphere brain damage, there is also some variability among the patients with right-hemisphere brain damage (Winner et al., 1998, pg 100, 101).

Winner et al. (1998) assume that ironic jokes are made to cover up ones embarrassment in case you are asked an ironic question. But I would argue that we are more likely to answer an ironic statement with irony to indicate

that we have understood the irony. This would be in line with Gibbs (2000) findings, their study investigated in irony in a natural language environment. They often found that ironic utterances are answered by the hearer with another ironic utterance (21 to 33 percent, depending on the type of irony), or laughter (12 to 25 percent) (Gibbs, 2000, pg 23). This is interesting, if we remember the grounding process involving three steps, there are at least two of them found here. As Gibbs notes, this shows how much irony is "a state of mind jointly created by speakers and listeners [...]. These ironic conceptualizations are often part of speakers' and listeners' common ground such that people will create ironic routines to exploit, and indeed celebrate, their mutual recognition of life's ironies" (Gibbs, 2000, pg 25). I will come back to this study in the discussion in (3.4).

Monetta et al. 2009 Monetta et al. (2009) conducted a study on irony comprehension among patients suffering from Parkinson's disease. The hypothesis reads as follows: What distinguishes irony from a lie is the assignment by the hearer of second-order beliefs to the speaker. The same is the case for some jokes. They tested this assignment of second-order beliefs with people having Parkinson. Parkinson's disease is a neurodegenerative disorder that causes depletion of dopaminergic neurons in the basal ganglia, which is the area where extensive neural connectivity takes place. Parkinson is known not only to affect motor abilities but also to cause cognitive impairments such as dementia or to harm the processing of language due to working memory deficits. Participants of this study were compared to a control group and tested for dementia, there was no difference between the groups, which suggests that the disease wasn't very advanced. Also the Parkinson patients were tested for cognitive- and frontal-lobe-tests where they scored comparable with healthy controls, with the exception for verbal working memory (Monetta et al., 2009, pg 974). This is of course crucial because it is predicted that those working memory deficits also influence pragmatic processing where knowledge of the context and inferential processes are required. Only if one can incorporate the context and the situation he or she can eventually understand the ironic meaning because he or she is aware of the contradiction taking place somewhere in verbal or nonverbal language. If this fails, the hearer will construct a lie where irony was intended, since both are false statements but with different intentional communicative goals. One factor that is important to make out the difference is theory of mind. Second-order theory of mind is the ability to understand that someone else is able to attribute mental-states to someone else. Deficits in this ability are assumed to cause deficits in irony comprehension. Since I have to understand that some-

one else is not lying to me but wants me to understand him or her properly and to do that, my opposite knows (or at least assumes) that I know that I have to construct the non-literal meaning (Monetta et al., 2009, pg 972, 973).

In this study non-demented Parkinson's patients were tested whether they can distinguish between lies and irony and if they can attribute mental states to others. They used the same lie- and irony-stories than Winner et al. (1998), which differ only to the regard that in one the protagonist knows his opposite would detect her story as false and in the other condition she does not know he knows and can tell that she is lying. Furthermore, what is interesting is that the stories were read in a neutral tone that means there were no special prosodic cues available. They found the healthy controls and Parkinson patients to understand the literal content of the story. But the Parkinson patients were significantly less accurate in answering the second-order believe question that verified the understanding of her knowing that her lie is obvious or not. Patients were also less accurate on the first-order believe question. Finally, Parkinson patients were less accurate when asked to judge whether the woman in the story was joking or lying (Monetta et al., 2009, pg 974, 975).

They conclude that pragmatic inferences like those necessary for irony comprehension are correlated with second-order theory of mind (Monetta et al., 2009, pg 978). That working memory and the processing of pragmatic meaning are correlated cannot be told with certainty by this study, a bigger sample than 11 patients would be needed to do so. What the frontal lobe does, which is particularly impaired among the patients of Parkinson's disease, is to coordinate so-called **executive functions**, for example affect organization and planning. Recent models of the interpretation of non-literal language that have been discussed above in (2.3) propose that the literal and the non-literal meaning have to be computed. So, the literal meaning is needed but has to be suppressed in favor of the non-literal intended ironic meaning (Monetta et al., 2009, pg 972, 979). "These processes for inhibiting contextually irrelevant meanings are likely to require greater mental flexibility to activate the nonliteral interpretation in context, which again exemplifies the intimate relationship between frontal lobe functions and many pragmatic aspects of language" (Monetta et al., 2009, pg 979–980). Other studies have shown involvement of the left medial prefrontal cortex, the right temporal lobe, and the medial orbitofrontal cortex for the ability to attribute mental-states to others and others have shown these regions to be involved in irony processing, as well (Monetta et al., 2009, pg 980).

So it is difficult to tell, what exactly caused the Parkinson's patients to perform worse than controls, whether it is the lack of verbal working

memory resources or the impaired executive functions or both. But the task, to differentiate between lies and jokes, seems to rely on the ability to derive inferences about the intentional communicative goals of the speaker.

McDonald 2000 To sum up this section, the following paper on neuropsychological studies of sarcasm is a review itself. McDonald (2000) focuses on the deficits caused by damage in the right hemisphere (RHD) and traumatic brain injuries in the comprehension of sarcasm relative to healthy controls and asks what these impairments can tell us about the types of inferences generated in sarcasm.

The right hemisphere has been shown to have a specialized role in the acoustic modulation of language, nonverbal functions like facial recognition, and "broader aspects of pragmatic language skills" (McDonald, 2000, pg 87). Also patients with right hemisphere damage have been shown to have problems with drawing inferences in language comprehension as is necessary for the use of metaphors, proverbs, idiomatic phrases, abstract relations between words, and the punch line of jokes. Patients had difficulties integrating a seemingly incongruent comment where it would have been necessary to grasp the emotional mood of an utterance, that is, make **emotional inferences**.

It also was shown for right hemisphere brain-damaged patients to have problems integrating the information about the relationship between speaker and hearer to understand whether a non-literal counter-factual comment is supposed to be either a joke or sarcasm. McDonald (2000) concludes that it is the emotional attitude that is a central part of a sarcastic utterance. Which is something the echoic mention theory predicts. If sarcasm is related to negative affection and the right hemisphere is involved in processing of emotional states, it is plausible that they fail more often in interpreting sarcasm. But it is not only emotion that makes the comprehension of sarcasm difficult for patients with damage in the right-hemisphere, as was shown above with the study by Winner et al. (1998), where patients failed to distinguish lies from jokes because they could not grasp the who-knows-what-pattern in a certain situation, in other words, a lack of (second-order) theory of mind. All in all, it depends on which theoretical framework one employs, to explain the deficits found. The line of reasoning McDonald (2000) choses, reads as follows. The right hemisphere seems to be involved in the processing of emotional aspects linked to language and they have been proven to be necessary to interpret sarcasm as the kind of sarcasm associated with negative emotion or attitude. However, as McDonald (2000) notes, it is also possible that the relatively poor competence of the brain-damaged patients is simply due to a more general loss of efficiency.

Traumatic brain injuries occur most often after an accident. A common impairment is damage to the temporal and frontal lobe that often causes a loss of communicative skills even though language itself is not impaired, as is the case for aphasia (McDonald, 2000, pg 90). For language production, they fail to meet social communicative goals, for comprehension, they only grasp the most concrete literal and salient meaning. That means they have difficulties with the social, pragmatic aspect of communication. For example, patients have shown to have difficulty making sense of an exchange like the following (McDonald, 2000, pg 91):

- (48) A: *What a great football game.*
B: *Sorry I made you come.*

Healthy speakers tend to interpret the first sentence of A as sarcastic but patients with a traumatic brain injury are not able to make the step of reinterpretation after reading the answer of B. Interestingly, even when the patients were listening to the conversation and were not impaired in recognizing the emotional tone of voice they still were poor in understanding the irony. This is another indication that prosody might be used as an additional clue for irony but is probably neither a necessary nor a sufficient feature. But it was associated with poor performance in abstract reasoning, concept formation, and flexibility of thought (McDonald, 2000, pg 91). They made out two different kinds of inferences necessary for the comprehension of irony: While in the example with the football game, the literally contradictory answer also includes the attitude of the comment, there are cases where there is no emotional judgment involved. For example (McDonald, 2000, pg 94):

- (49) A: *That's a big dog.*
B: *Yes, it's a miniature poodle.*

This is a simple ironic contradiction without involvement of any attitude of the speaker, besides that he or she wants to make fun of the fact that the dog is very small. The patients performed differently for the two cases.

"In [the example in (48)], the participants had to infer that the first speaker was being critical because the second speaker was apologizing. They then had to make a subsequent inference that the first speaker thought the opposite to what they stated. Thus, there were at least two necessary steps to the inferential process. In [the example in (49)], the only inference necessary was that the first premise was being responded to as if it meant the opposite to that literally asserted" (McDonald, 2000, pg 94).

Another possible explanation for the patients better performance for the latter kind of irony is that simple contradiction might be more obvious, whereas the former example requires to make inferences about the sarcastic persons intentions. This is an aspect of inferential reasoning that has to do with the illocutionary force of an utterance, or in other words, with what the speaker wants from the hearer. Which poses neuropsychological evidence for the hypothesis that ironic meaning has to be inferred as reciprocal process between the interlocutors, where not all inferences are equally easy or hard to draw. It also supports the definition of irony, where sarcasm is understood as subclass of irony that is different because here an emotional attitude plays a role, whereas this is not the case for all instances of irony. It has been shown in resulting in different levels of difficulty for impaired patients suffering traumatic brain injuries (McDonald, 2000, pg 94,95).

Summary There is clear evidence that theory of mind is essential to understand irony and sarcasm. What has also been shown, is that theory of mind comes in different variants: first- and second-order. Maybe the involvement of theory of mind differs for different kinds of sarcasm, depending on how subtle and how multilayered the speaker's intentions are. Also, it depends on the theoretical framework the researcher employs, to explain the deficits that were found. If one follows the echoic mention theory in the relevance theoretic framework, one is likely to argue with the processing difficulty being caused by the emotional meaning conveyed. If one follows the graded salience hypothesis one wants to explain processing difficulties with a two-step interpretation and more processing load since the literal and non-literal interpretation have to be kept in mind to compute the mismatch between the two.

What is sure: For irony a second-order theory of mind is needed. This ability to draw inferences about the mental state of the other person is necessary to make predictions about what the speaker wants to communicate. So it comes down to emotional inferences to understand communicative intentions. These communicative intentions will be the subject of the next section.

3.3 Communicative intentions

What kinds of mental states do we attribute to others using our theory of mind? We can detect pretense and that means we draw pragmatic inferences and predictions about the relevant meaning. In other words, we reconstruct communicative intentions which is crucial to detect the underlying meaning

of irony. This is the last argumentative step I would like take while discussing psycholinguistic literature.

Creusere 2000 Creusere (2000) was also interested in children's recognition of irony and the communicative function conveyed in irony. As already mentioned, for the echoic mention theory the main communicative function of irony is to indicate the speaker's attitude towards the situation and the persons involved. There has been other evidence, though, against the echoic mention theory, that a central function of irony is humor. The allusional pretence view of irony claims that irony derives humor and mutes criticism or praise or whatever attitude one literally expresses. Children have been shown to understand such kinds of communicative functions even when they did not grasp the ironic meaning itself (Creusere, 2000, pg 31, 32).

Her study aimed to shed some light on the communicative function of irony and its acquisition. She asked, whether 8-year-old children are able to understand what is necessary for irony, that is, to communicate allusion and pragmatic insincerity (Creusere, 2000, pg 34).

She wanted to test the allusional pretence view of irony, a theory that assumes the opposition-definition of irony is too narrow. Nevertheless, she admits that contradictory forms, that follow the classical definition of irony as opposition are the prototypical cases of irony and therefore should be understood more easily than, what she categorizes as irony in true assertions, offerings, and expressions of appreciation (Creusere, 2000, pg 35). Just like the following examples of pragmatic insincerity and allusion:

- (a) true assertions, such as "You sure are hungry" to a person who just ate half of a pizza meant to be shared among five people; (b) over-polite requests, such as "I hate to bother you, but would it put you out too terribly much if you refrained from walking naked in front of your living room window" from a neighbor with kids who frequently play in their front yard across the window; (c) questions, such as "Would you like another beer?" to a guest who apparently already had enough to drink and was becoming obnoxious; and (d) offerings, such as "Here, warm up with a few practice balls" to a bowling opponent who had just thrown three strikes in a row (Creusere, 2000, pg 33).

She tested what participants considered as ironic. Here an example of a scenario:

One day Sue was sitting in the kitchen talking to a friend on the phone. Bill came into the kitchen and started taking a bunch of

pots out of the cabinets. Every time he put a pot down, it made a loud clanging noise. Sue had to raise her voice so that her friend could hear her on the phone. Finally, Sue got flustered, told her friend "bye" and hung up the phone. When she turned toward Bill, he said good morning. Sue said to Bill [either Version A, B, C, D, or E]:

A. Counterfactual assertion: "Oh hi. I didnt even notice you were here."

B. True assertion: "Oh hi. You know, I couldnt help noticing already that you were here."

C. Question: "Oh hi. Did you notice that I was trying to talk on the phone a second ago?"

D. Offering: "Oh hi. Heres another pot. You can bang it too!"

E. Thanking: "Thanks a lot for banging the pots while I was on the phone (Creusere, 2000, pg 36).

Children had to answer a fact, memory, purpose, meaning, and sincerity question. (*Was Bill quiet or did he make a lot of noise with the pots? What did Sue say to Bill? Why did Sue say that to Bill? What did Sue mean when she said that to Bill? Did Sue appreciate Bill banging the pots?* (Creusere, 2000, pg 36-37))

So children had to rate the humor and meanness of the possible answers. Also they were asked to give first-order-believe-judgments ("Did Sue think that Bill was quiet or noisy?" (Creusere, 2000, pg 37)) and second-order-believe-judgments ("Did Sue want Bill to know what she was thinking?" (Creusere, 2000, pg 37)). That were the meta-linguistic questions. Children who answered first- and second-order questions right were coded as understanding the irony.

The results show that children were able to detect the allusion in those utterances and the pragmatically insincere component. Furthermore, the data suggests that comprehension of allusion and pragmatic insincerity are related to the first- and second-order believes measurement. That means the two methodological ways to measure irony are not independent of each other (Creusere, 2000, pg 37). The prediction that counterfactual statements are more easily understood than offerings, true assertions, questions, and expressions of thank was not confirmed (Creusere, 2000, pg 38). The question was rated the most sincere, than they rated the counterfactual and offering forms of irony and the most insincere were the true assertions and the thanks-form.

The thanking in E. received the lowest "funniness" score but the highest "meanness" score. The author concludes that the prediction that there is an

effect of propositional form on humor and meanness was only partially confirmed. Counterfactual instances of irony were not interpreted as particularly funny compared to the other utterances. Furthermore, apparently children at the age of 8 are not yet able to interpret an utterance as humorous and mean at the same time. This is in line with other studies that show that the sensitivity to the humor function of irony increases over age (Creusere, 2000, pg 41). All of the children responded correctly to the questions that tested the understanding of social conventions. There cannot be drawn a line from propositional form of the utterance and the comprehension of irony. For the children it was especially hard to understand the pragmatic insincerity in the question in C. I would argue that this is difficult to understand as a rhetoric question for adults, too. The form of the utterance did not facilitate the recognition of first- and second-order believes. Neither did it clearly have an effect on the children's ratings of meanness and funniness: There was no main effect of form for speaker funniness, but there was one for form and speaker meanness (Creusere, 2000, pg 42).

What the two studies on children's acquisition of irony that were so far described have in common is that they propose a general skill that makes children able to understand the communicative functions of irony and propose that there is a general skill that allows to understand irony but does not let them yet distinguish between the fine-grained forms of ironic speech.

Cheang and Pell 2006 Cheang and Pell (2006) write about humor and communicative intention and how they might be impaired for stroke victims with right hemisphere damage (RBD). They argue that understanding humor requires to decipher the underlying implicit meaning which has to be derived from the **communicative intention**. To them communicative intention is the implicit message conveyed in an humorous utterance. Their example:

- (50) a. I could care less.
- b. I couldn't care less.
- c. I don't care at all.

According to Cheang and Pell (2006, pg 448) in popular language use, they all mean the same thing. Which makes a. an ironic phrase, or at least it used to be ironic and now is a frozen ironic utterance that we could maybe call a joke or a saying, where the speaker's meaning expressed is unconcern. (This might be an example of when irony can become conventional in the Gricean sense.) They report joke completion studies where RBD participants were found to sort the punchline to jokes. They performed significantly worse than controls, but they did not make completely arbitrary mistakes. Rather,

they consistently chose the wrong ending that were surprising but unrelated (slapstick ending). That means RBD patients are aware of the fact that a punchline should be surprising, but they were not able to construct a background that would make the text of the joke coherent (Cheang and Pell, 2006, pg 448-449). This is very interesting since it is exactly the same finding than what Sabbagh (1999, pg 51) writes for ASD participants, that will be discussed in the next section. They too did know that the punchline of a joke is supposed to be surprising but failed in choosing the consistent ending that required reinterpretation of an absurd context that would pose the right background for the joke to make logically sense.

For RBD participants it is reported that the interpretation of sarcasm is especially difficult (Cheang and Pell, 2006, pg 449). They let a RBD group and a control group take a joke and story completion task and a task to access their ability to judge pragmatic intent. The story completion task was to check participants ability to integrate information in narratives that did not contain surprising elements while the joke completion task did contain such a surprising element. An example story reads like this (Cheang and Pell, 2006, pg 452):

Mable walked into a pastry shop. After surveying all the pastries, she decided on a chocolate pie. ‘I’ll take that one’ Mable said to the attendant, ‘the whole thing’. ‘Shall I cut it into four or eight pieces?’ the attendant asked.

The following endings were possible:

- (51) a. “Well, there are five people for dessert tonight, so eight pieces will be about right” = Story completion ending
- b. Mable said, “You make the most delicious rolls in town” = Unrelated ending
- c. “Then the attendant squirted whipped cream in Mable’s face.” = Slapstick ending
- d. Mable said, “Four pieces, please; I’m on a diet” = Joke ending

So the answer in d. was the right one. For the other condition, the story completion task, the last sentence was replaced with:

- (52) “My son just graduated from law school” = non-sequitur ending

The second task was a pragmatic interpretation task, where depending on the relationship between actor and speaker literally false contexts could be sarcasm, jokes, or lies (Cheang and Pell, 2006, pg 453). The results for

the joke completion task compared to story completion were not clear cut. Even though there is a tendency that indicates for RBD patients the joke completion was more difficult there were also RBD patients that scored like controls (Cheang and Pell, 2006, pg 455). For the pragmatic interpretation, however, there was an effect but not between the groups. Both patients and controls were more likely to judge the non-literal comment as lies (to spare the actor's feelings) when their relationship was friendly. When they were characterized as enemies, the participants of both groups were more likely to infer a sarcastic meaning. Still, RBD patients were significantly less likely to interpret a comment as a joke when speaker and actor were friends compared to the control group (Cheang and Pell, 2006, pg 456). The authors interpret this as maybe indicating that RBD patients had problems integrating information about the relationship between the interlocutors and having difficulties with the subtleness of conversation (Cheang and Pell, 2006, pg 459).

There were no clear results in this study, even though the trend indicates that RBD patients might be impaired in appreciating humor, which would be consistent with previous findings. I still think the study hints on interesting points: First, relationship information might be important to interpret jokes, irony, and lies. Second, the difference between an arbitrary surprise and a "real" joke ending is important. A joke ending cannot be completely arbitrary, but follows its logic, integrated in an absurd context. We will come back to this point in the discussion in (4) and in (5), since it also might be a feature for an ironic meaning.

3.3.1 Communicative intentions and Autism

As I mentioned at the beginning of the section about theory of mind in (3.2), patients within the autistic spectrum disorder (ASD) have played a big role in the empirical analysis of pragmatic comprehension, since they are said to have problems with especially this aspect of language: They are impaired on verbal and nonverbal communication due to a lack of social competence (Baron-Cohen et al., 1985, pg 37-38). Describing this fairly general diagnosis more precisely has been a main goal of linguistic research among ASD children during their language acquisition. Since Autism in its different variations can never be described by the same impairments I will always write of children or adults ranging within the autistic spectrum disorders, indicating that they share common features of impairment, and in most cases for language studies only high functioning individuals are participating, but one cannot be mistaken and assume a fairly homogeneous impairment. Rather, they differ widely.

Bearing in mind the basic claim by Baron-Cohen et al. (1985, pg 43) for children in the autistic spectrum disorder, which was that what they lack precisely for successful communication is a theory of mind, I will now discuss some examples of the line of psycholinguistic research that is interested in how autistic people understand irony. Most studies are especially investigating in the understanding of sarcasm and its communicative intention.

Wang et al. 2006 The following study by Wang et al. (2006) explains the impairment among children with autism spectrum disorders with the failure to understand contextual clues, such as sentence context, global inferences, and especially prosodic cues. It is known that children with ASD are impaired in extracting emotional meaning from tone of voice (Wang et al., 2006, pg 933). In their fMRI-study they focused on the neural circuitry supporting higher-level pragmatic processing such as irony comprehension, where even high functioning adults with ASD are impaired. Previous studies have observed abnormalities in the medial prefrontal cortex (MPFC) relative to controls. It was found, though, that if ASD children were instructed to pay attention to the facial expressions and tone of voice of the speaker the activity in MPFC increased, suggesting that neural functioning it is intact. They suggest that different cues, such as facial cues, intonation, or context, can replace each other, meaning, if one is very strong and clear cut, irony can function with only one cue. What was expected was poor performance of children with ASD compared to controls in detecting irony when only one clue was given, either intonational or contextual information. Also less activity in brain regions known to be involved in theory of mind abilities was expected (Wang et al., 2006, pg 934).

The participating children (the age of the ASD children was 7.4-16.9 years, the typically developing children were 8.1-15.7 years of age) all performed above chance in detecting irony, with the typically developing children being significantly more accurate. The typically developing children performed more accurately than children with ASD where contextual information was provided. When only prosodic cues were available the groups did not differ. Significant activity in the MPFC was only found for the typically developing children. ASD children showed higher activity in the right inferior frontal gyrus, in the temporal regions bilaterally, and in the left pre-central gyrus (Wang et al., 2006, pg 937).

Expected was poorer performance of the ASD children no matter if contextual, or prosodic information was given, or both, because ASD is said to be associated with problems with incorporating contextual information and extracting emotion from tone of voice (Wang et al., 2006, pg 934). In

sum, they found differences between typically developing and ASD children that indicate difficulties for the ADS children when it comes to incorporating contextual information to infer speaker's intent. The groups did not differ significantly if only prosodic cues were given (Wang et al., 2006, pg 937). They explain the non-existing performance difficulties between groups when there was only intonational information with greater difficulties for the typically developing children and not with possible strength of the ASD children (Wang et al., 2006, pg 938). This is further evidence for the claim that prosodics only plays a minor role in comprehending irony. So maybe, where typically developing children combine different clues, such as context and intonation, this task of exploiting different clues to detect the hint to construct an ironic meaning is what is difficult for ASD children and therefore there is no performance difference between TD and ASD children when there is only one clue and not different hints that have to be combined to retract the necessary information.

In terms of brain regions, ASD children showed stronger activity in prefrontal and temporal regions than typically developing children. The regions they activated more strongly were the same the typically developing children used. That may be relative to increased task difficulty and reflect problems in the integration of contextual cues to interpret speaker's intention. More activation might be due to greater task difficulty for the ASD children. The ASD children that had better social communicative skills showed more activation in the temporal poles. This is contrary to what was expected. Expected was poorer brain activity in the regions associated with mentalizing since ASD children are said to have problems with theory of mind (Wang et al., 2006, pg 934). This might be due to the fact that children were explicitly asked to judge the communicative intension, while in previous studies this was implicitly required to fulfill the task. This means ASD children can activate the MPFC when when the task explicitly requires them to interpret mental states of others. This suggests a difference in implicit and automatic processing. ASD children are able to bring the relevant brain region on-line if explicitly instructed, that is, when task demands attention to mental states (Wang et al., 2006, pg 940). That the ASD children with better language abilities were more likely to show higher activity can reflect compensatory strategies that replace normally instinctive processes of interpreting communicative intention. So the ASD children were able to interpret sarcasm, but it comes at the cost of more computational effort (Wang et al., 2006, pg 939-940).

Sabbagh 1999 On communicative intentions, Sabbagh (1999, pg 29) again points out to the specific language deficit that ASD individuals seem to have compared to patients with RHD and calls them “extra-linguistic” problems, including deficits with prosody, non-literal speech, and deciding what the main theme in discourse is and hence making appropriate contributions accordingly. For children those problems can be summarized as integrating information in its context. Sabbagh (1999, pg 30) wants to answer the question, what context is by reviewing studies made with patients with those impairments. He proposes as an answer “an ability to make appropriate inferences about one’s conversational partner’s *communicative intentions* (CIs)” (Sabbagh, 1999, pg 30). To make inferences about an interlocutors CIs it is essential to have a common ground, that is, an understanding of what is mutually known. This means one has to have a theory of mind to know what the common ground is and from that background one can derive the communicative intentions of one’s opposite. This means, we are back at a dynamic model of a common ground that is updated within discourse. Sabbagh (1999, pg 32) calls that “discourse tailoring”. For humor this especially is the case: “expressions that are at once incongruous with a given conversational context, but can be resolved after a reinterpretation of the context is made” (Sabbagh, 1999, pg 50). This recalls the point made at the end of the discussion of theories: that irony alters the context in a way that challenges us to construct a ridiculous background which is what amuses us. Or the figure ground reversal that we have seen for jokes. RHD patients according to Sabbagh (1999, pg 51) had problems with reinterpreting the context to yield the humorous interpretation. ASD individuals seemed to understand that the punchlines of jokes should be surprising but they failed in choosing the endings that were surprising **and** “resolvable within the context of the story” (Sabbagh, 1999, pg 51). That means they were not able to achieve a reading that was **logical** in front of the background that was to construct.

Happé 1993 One study Sabbagh (1999) mentions, is the one by Happé (1993). Since it is one of the earlier ones I want to have a look at it more closely. She tested communicative competence, theory of mind, and relevance theory. This is about constructing another person’s false believes: Irony demands second-order metarepresentations **joint attention** (Happé, 1993, pg 102). As relevance theory claims, being ironic means someone is referring to a previous thought or utterance. Therefore it is different to the processing of metaphors, which require only a first-order theory of mind. This supposedly is a different prediction as the predictions made by classical theories (Happé, 1993, pg 104). The paper by Reimer (2013) that I reviewed before in chapter

(3.2) argues against this claim that Gricean theories predict the same amount of processing for irony than the processing of metaphor requires. The findings by Happé (1993) suggest, as expected within the relevance theoretic theory, a difference between the processing of metaphor and irony. She distinguished between No-ToM Autistics, 1st-ToM Autistics, and 2nd-ToM Autistics. For metaphor the groups with No-TOM did not perform different while for irony there was a difference in performance between the 1st- and 2nd-ToM groups (Happé, 1993, pg 111). The same was the case for healthy children that she classified whether they had already developed a second-order theory of mind, or not. The ones with second-order theory of mind performed nearly perfect on irony, while failers performed worse. For metaphor comprehension there was no difference in performance between the groups (Happé, 1993, pg 113). She concludes that second-order theory of mind is a good predictor of comprehension of irony and this is in line only with relevance theory but not with Gricean accounts (Happé, 1993, pg 115). She summarizes:

Whatever distinguishes the autistic subjects in the three theory of mind groups has a direct and particular association with the comprehension of figurative language. In the absence of other differences, the confirmation of the predictions made would seem to suggest a real difference among the groups in underlying theory of mind competence (Happé, 1993, pg 116).

So the classification in groups with second-order or first-order theory of mind is crucial. However, I would like to point out to the material that is listed in the appendix as an example for the metaphor versus irony task. She uses the following “cake story” (Happé, 1993, pg 119):

David is helping his mother make a cake. She leaves him to add the eggs to the flour and sugar. But silly David doesn't break the eggs first - he just puts them in the bowl, shells and all! What a silly thing to do! When mother comes back and sees what David has done, she says: “Your head is made out of wood!”

Q: What does David's mother mean? Does she mean David is clever or silly?

Just then the father comes in. He sees what David has done and he says: “What a clever boy you are, David!”

Q: What does David's father mean? Does he mean David is clever or silly?

The right answer is primed because the word “silly” is mentioned in the text two times. I would think that to chose this answer, there is neither comprehension of irony, nor metaphor necessary.

Frith and Happé 1994 Frith and Happé (1994) argue in another paper that language isn't the same as communication. "Language is a grammar-governed representational system" while "communication is a process in which one person alters the physical environment of another in such a way that the other constructs internal representations similar to those in the head of the first" (Frith and Happé, 1994, pg 97). Obviously the preference for intentions over words is very clear for non-verbal communication. Also children seem to be more interested in intentions than words, too. They conclude that during language acquisition there is a "precedence of communication over language" (Frith and Happé, 1994, pg 97). Bringing this together with the hypothesis that children and adults with a disorder in the autistic spectrum lack a theory of mind, this relates closely to the ability to understand non-literal speech such as metaphor, similes (metaphors plus "like" such as "you dance like a swan", compared to "the dancer was a swan"). This brings us back to joint attention that was mentioned already in (3.2), this plays a huge role in the acquisition of vocabulary, that seems to be delayed for ASD children (Frith and Happé, 1994, pg 100). When the mother is talking about something it is essential that the child understands about what she is talking, that is called joint attention, and it obviously has to do with the child's ability to put himself or herself in the mothers place. This means that language acquisition relies on communicational skills (Frith and Happé, 1994, pg 102).

McDuffie et al. 2006 This leads to a last argumentative step, before we will come back to irony: McDuffie et al. (2006, pg 423) make the same point when they discuss fast-mapping among ASD children. "Gaze-monitoring", "joint attention", or "attention-following" mean all the same thing. That is the basis for forming associative pairs between a heard label and the object of current attention that is called **fast mapping** (McDuffie et al., 2006, pg 422). If the ability of ASD children for word-object mapping is impaired due to a deficit in attention-following, because they "incorrectly used their own focus of attention as a strategy for making word-object associations rather than using the speaker's gaze direction to guide their word learning" (McDuffie et al., 2006, pg 424). If this is the case, the same mechanism takes place than for irony comprehension. The children were presented with verbal and nonverbal cues to support attention-following within a certain context of social interaction (McDuffie et al., 2006, pg 428). As an outcome attention-following seemed to be related to fast-mapping that itself was related to vocabulary outcomes (McDuffie et al., 2006, pg 431). This is interesting for comparison to ASD and irony because when more explicit cues were provided,

the children with ASD improved (McDuffie et al., 2006, pg 433), just as they understand irony better when directly pointed to it, as we saw with the study of Wang et al. (2006). For this study one has to mention that there was no control group, as they mention themselves. And for a design that makes the claim to pose insight in regular vocabulary learning, it would be interesting to see if their hypotheses hold for unimpaired controls, as well.

Summary What I wanted to show in this last section is that there are general skills one can relate the comprehension of irony among ASD children to. This clearly leads away from the core issue about irony comprehension, but it shows how general the skills are that come into play here. Research on ASD reflects the issues discussed before in chapter (3.2): Theory of mind is necessary to predict communicative intentions of the speaker. ASD participants are not unable to understand irony, but they need to be pointed to it. So the complexity of automatically understanding that there is an implicit meaning that has to be derived by using different clues seems to be too much for them.

3.4 Discussion

Summarizing the discussion above, I will point out to two studies that make more general claims about irony and were therefore excluded from the argumentation above. But bringing this all together I think they provide fruitful insights.

Gibbs 2000 Even though Gibbs' understanding of irony is much wider than the definition adapted in this thesis, and his perspective on irony is a completely different one, his study is interesting since he focuses on the social purposes irony fulfills and not on its structure. And that is the main point I would like to focus on in this discussion of psycho- and neurolinguistic research. Gibbs (2000) makes clear that jocular language is a means of communication that is very often employed in natural dialogue (among friends on a college campus) and therefore he makes some claims about the communicative function of irony.

Incorporating in his definition of irony he names jocularity, sarcasm, hyperbole, rhetorical question, and understatement. This means, he also counted instances of speech where the speaker was communicating something less than what he or she was saying as ironic (Gibbs, 2000, pg 12,13). Counting all those as instances of ironic speech he finds that it makes up eight percent of the turns in 62 conversations among college students (Gibbs, 2000,

pg 14). He assumes, thus, that ironic talk can have many different social functions:

Irony is routinely used in the ongoing flow of conversation between group members to affirm their solidarity by directing comments at individuals who are not group members and not deemed worthy of group membership. [...] Sarcasm, in particular, is often used to vent frustration when an individual finds some situation or object offensive or sees a group's normative standards violated (Gibbs, 2000, pg 7).

Facing this huge amount of different forms irony can inherent, he doubts the possibility to formulate one unifying account on how irony functions. He investigated sixty-two five to ten minute conversations. Finding that, as is often assumed in the literature that jocularity, sarcasm, and hyperbole, as the most common cases of irony, mainly refer to people. Even when someone is talking in an ironic way about a thing the comment is transferred to the related person that owns it or has something to do with it (Gibbs, 2000, pg 15, 18, 19). This fits to the assumption employed in this thesis that, in contrast to cynicism, irony and sarcasm relate closely to an issue or thing touched upon in the present situation and what could be more salient than what is related to the two interlocutors or other people they talk about. 50 percent of the ironic instances were jocular language, 28 percent were sarcasm, 12 percent hyperbole, eight percent rhetorical questions, and two percent cases of understatement (Gibbs, 2000, pg 15). He defines jocularity as "where speakers teased one another in humorous ways" and sarcasm as "where speakers spoke positively to convey a more negative intent" (Gibbs, 2000, 12).

Gibbs checked the jocular and sarcastic utterances whether they are conveying a negative meaning by stating something positive (saying: *You are a fine friend* while meaning *You are a bad friend*) or whether they are literally stating something negative to express a positive meaning (saying: *You are a bad friend* while meaning *You are a good friend*) (Gibbs, 2000, pg 14). He found for jocularity that in nine percent of the cases a positive statement was uttered to express something negative. 28 percent of the cases were negative statements uttered to convey a positive meaning and in 62 percent he could not tell, or they were not expressing a meaning via literal opposition. Unfortunately, he does not give any examples of these instances, neither does he say what cases of jocularity they were if they were still classified as ironic. Gibbs concludes from this finding that approaches that assume irony to express the opposite of a literal meaning do not come close to grasping the complexity of

ironic language use (Gibbs, 2000, pg 17). For sarcasm, he found 69 percent of cases having a negative meaning while something positive is said and 15 percent were negative statements meaning something positive. For 17 percent he could not classify them or they did not express an opposition between meaning and literal utterance (Gibbs, 2000, pg 19).

What Gibbs finds about special intonation for irony (for example: nasalization, slow speaking rate, exaggerated stress on certain words (Gibbs, 2000, pg 10)) is that it is present in most of the cases, but not necessary for irony, since there are also cases without any special intonation patterns. However, it can serve as a clue even though there is not one special prosodic pattern for ironic speech (Gibbs, 2000, pg 25).

He often finds that ironic utterances are answered by the hearer with another ironic utterance (21 to 33 percent, depending on the type of irony), or laughter (12 to 25 percent) (Gibbs, 2000, pg 23). This is interesting, if we remember the grounding process involving three steps, there are at least two of them found here. As Gibbs notes, this shows how much irony is "a state of mind jointly created by speakers and listeners [...]. These ironic conceptualizations are often part of speakers' and listeners' common ground such that people will create ironic routines to exploit, and indeed celebrate, their mutual recognition of life's ironies." (Gibbs, 2000, pg 25).

His results show that not all instances of irony are a case of echoic mention, pretense, or allusional reminders as for every one of his categories of ironic use there were cases without any echoic mention or pretense (for jocular language there was the most echoic mention, while in general pretense was more often involved than echoic mention). "Yet, it seems clear that almost every ironic utterance in the corpus examined here alluded to speakers' and listeners' expectations and indirectly conveyed speaker's attitudes through some form of pragmatic insincerity" (Gibbs, 2000, pg 24).

In general, it is difficult to argue against Gibbs' point of irony not being just the opposite of the literal utterance, without knowing what the cases looked like that were classified as ironic but not expressing literal opposition. Even though, Gibbs' definition of irony is too wide to say anything about the meaning of irony in particular, but what his study shows, is the importance of jocular speech for everyday conversation and how much it is used (at least, for young people among their friends in the late 1990s on a college campus, as was the population investigated here). However, the distinction between what he calls jocularity and sarcasm isn't clear to me, at all. He merely states that sarcastic speakers were more critical and mocking as the speakers of jocularity. The type of irony he calls understatement isn't even viewed as humorous, what, as to my concern, is the very basic feature of irony. He merely states: "Most notably here, there appears to be a strong association

between an ironic utterance mocking someone or something and it being viewed as humorous” (Gibbs, 2000, pg 23). Which does not say much about the meaning of irony, however, with his more broad view on irony he does find evidence for some of the theoretical postulates made like the three-step process of conversation. Also, he shows in natural dialog that speakers and listeners create meaning together by matching their expectations and predictions about what the opposite wants to communicate. Which is another way to describe a grounding process and how in doing that meaning is enriched.

A second study that I want to discuss here is the one by Filik et al. (2015), because it provides evidence for the point I made at the end of the theoretic discussion in (2.6.3), where I stated that irony might enrich meaning not by strengthening but weakening.

Filik et al. 2015 conducted a study about sarcasm and emoticons with interesting results. First, one has to mention that they define sarcasm quite differently as is defined in this thesis. Sarcasm is defined as a special form of irony that always targets persons, either in a criticizing or a praising way. This corresponds to how in this thesis irony and sarcasm are defined, irony being what they understand to be praising sarcasm. So, when reviewing their study I talk of sarcasm meaning both irony and sarcasm. They claim that written sarcasm might be difficult to detect, since non-verbal clues like facial expressions or tone of voice might be missing. Their hypothesis is that emoticons might be used as aids to make the sarcasm easier detectable. Therefore they investigated in criticism or praise communicated literally or sarcastically and how emoticons or punctuation marks might be used in an ambiguous or unambiguous context (Filik et al., 2015, pg 1). An emoticon is defined as a glyph that uses signs out of the American Standard Code for Information Interchange (ASCII) in order to communicate an emotional state. In doing so they also wanted to provide an answer to the question, why do we use sarcasm at all, if it comes with the risk of misunderstanding. Most research claims that sarcasm also communicates some kind of social or emotional function that makes worth the risk. They aim to find evidence for the so-called **tinge hypothesis**, which assumes that sarcasm reduces the emotional value of praise and criticism. Sarcastic praise becomes less positive because it is ”tinged” with the literal negative meaning and sarcastic criticism becomes less negative since it is ”tinged” with the positive literal interpretation (Filik et al., 2015, pg 2).

I have a very strong intuition about the use of :) and ;) for humorous written language. :) communicates that something is meant not really seriously but rather as a joke, but leaves the literal interpretation of the sentence

unimpaired. This emoticon clearly communicates something like *I am kidding so do not take this too seriously, but there is a truth to that*. As it is most of the time the case, for jokes at someones expense, that they communicate some sort of truth but in a blurred or tinged way. Just like the jokes I described in reviewing the general theory of verbal humor, what the joke about Polish people actually makes fun of, is the stereotypes and prejudices we really have. ;) on the other hand, clearly communicates to the receiver that he or she needs to reverse the literal interpretation in order to derive the intended meaning, like saying *read the opposite*. Again, what amuses us is surprise, this can be about the actual truth in something obviously ridiculous (this is the case for jokes), or the untruth about something that could just as well have been true (this is the case for irony).

In a first experiment Filik et al. tested the effects of the emoticons :-P, ;-), compared to the punctuation ... and ! (Filik et al., 2015, pg 4). Their results suggest that the tongue face emoticon (:-P) and the wink emoticon (;-)) mark sarcasm and increase it, but only if the utterance is already perceived as sarcastic, that means there is a mismatch between the context and the utterance. It had no effect if the context suggested the literal interpretation. The two emoticons having no difference in the amount by which they increase sarcasm. The presence of punctuation had no effect in the literal or sarcastic comments. The emoticons had a significantly bigger effect on the literal criticism compared to punctuation (Filik et al., 2015, pg 6). That means, the wink and the tongue face emoticon intensified criticism either way, no matter if it was literal or sarcastic. On literal praise the emoticons and ellipsis (...) had similar effect. For the perceived effect of the comment emoticons make both, the literal and the sarcastic criticism, appear less negative. The two emoticons did not differ in their effect but compared to punctuation the effect was significantly bigger (Filik et al., 2015, pg 8). For literal praise, the two emoticons and ellipsis made literal praise less positive, while the exclamation mark made it more positive (Filik et al., 2015, pg 9).

The results provide positive evidence for the tinge hypothesis: with or without device (emoticon or punctuation) sarcastic praise was perceived as less positive than literal praise and sarcastic criticism was perceived as more positive than literal criticism (Filik et al., 2015, pg 9). Therefore they conducted a second experiment to investigate the meaning of these devices when the context does not provide much clues to understand an utterance as sarcastic. In the second experiment, thus, the utterances were ambiguous between literal and sarcastic interpretation. Since the tongue face had similar effects than the wink face in the first experiment, only the latter remained, also since the exclamation mark seemed not to have a huge effect on sarcasm, it was dropped. So only the effects of wink face and ellipsis were compared to a full

stop. They changed the question that aims at the perceived emotional effect, from "How ironic do you think the final comment is?" (Filik et al., 2015, pg 6) to "How sarcastic do you think Person A is being?" (Filik et al., 2015, pg 11), which is a little bit suspicious and not really justified from experiment 1 to experiment 2. As a result, positive comments were rated most sarcastic when accompanied by a wink emoticon, than by an ellipsis, and least sarcastic with a full stop. This is in line with research that suggests that wink emoticons make ambiguous comments more sarcastic, and also in line with my own intuition. In terms of emotional perception, a wink emoticon makes negative comments, like sarcastic praise and literal criticism, be perceived less negative, positive comments, like sarcastic criticism and literal praise, however, are neither perceived more positive, nor attenuated in their positiveness (sarcastic criticism was less negatively perceived accompanied by a full stop; sarcastic praise was more positively perceived when accompanied by a wink face; literal criticism was less negatively perceived when accompanied with a wink; and literal praise finally was perceived more positively accompanied by a full stop) (Filik et al., 2015, pg 11,12,13).

To sum up, in an ambiguous context only negative emotions are attenuated by a wink, while positive emotions are not altered either way by a winking emoticon. This supports my assumption of sarcasm as conveying negative comments. Compared to irony that does not have to convey sentiments towards the ironized issue and therefore there are no emotions that could be attenuated by a supportive device such as a wink emoticon.

What did we see so far? There is more evidence that supports a two-step-process of irony processing, than for a relevance-theoretic one-step account. Furthermore, it seems as if different clues are required to detect irony. And normally they are combined to detect the relevant information, which allows to construct the ironic meaning. Maybe the combination of different clues is what poses a problem for ASD children. Because they have been shown to be able to understand an ironic meaning if explicitly pointed to the task. But in a natural language environment communicative intentions are not that clear cut and sometimes they are a combination not only of different hints and clues but also the intentions themselves come in a graded fashion. One hints at something and the additional layer of meaning can, but must not be detected, like this meaning is created in a multileveled reciprocal way between the speaker and the listener.

There was evidence for ASD individuals (Sabbagh, 1999, pg 51) and RBD patients (Cheang and Pell, 2006, pg 448-449) that they both could not choose the correct punchline because they weren't able to construct an absurd but coherent context as background that would make the joke consistent. They

did choose surprising but completely unrelated endings. Which supports the view I discussed at the end of the chapter on theories about irony comprehension in (2.6.3), where I concluded that to construct the correct ironic meaning, one has to apply a ridiculous but logically consistent background as context, just like a figure ground reversal which was mentioned as a logical mechanism used for jokes. The non-literal ironic meaning stays the same, but the background changes to a ridiculous world which incorporates the non-literal meaning. This reminds of the theory that was just briefly mentioned in the beginning: the mental spaces theory, except that not the ironic utterance gets outsourced, but briefly we adopt a world view, where something ridiculous makes sense. Just like a temporal scenery against which the joke or the irony makes perfect sense. That is why good jokes are inherently consistent and in a ridiculous sense true, just like irony tells you something about the absurd world you have to imagine for it to be true and in contrast to that it ridicules the real world.

This is comparable to the study by Giora et al. (2000) where the mistakes by the RBD and LBD patients were telling. The patients that did understand the passage but not the irony chose "lie" incorrectly as an answer, which means they knew that the speaker had some intention in mind, which means they had a theory of mind, otherwise they would also, just like the participants that did not understand the passage, have chosen the "literal truth" answer (Giora et al., 2000, pg 73). So they also failed in constructing a background that would make out of a false statement an ironic one. Because without incorporating an ironic statement into a context that is ridiculous enough so that the ironic meaning makes sense, it is plainly a lie that remains.

Chapter 4

Situational and computational irony

In this chapter I intend am going to outline some common features of jokes and irony to then discuss some computational models on irony. To start right with an example. We consider it as ironic if we hear the story from a

(53) Management school that went bankrupt.

To show the similarity, an example from Attardo and Raskin (1991, pg 307)

(54) Gobi Desert Canoe Club

that they bring in as an example for their "logical mechanism[s]" (Attardo and Raskin, 1991, pg 306) in language.

It can be seen that the two are not that distinct at all. Jokes of this kind and what we consider as ironic in situations in real live follow the same logic. It is the last thing that we would expect a management school to fail due to a management error because this is the very thing they should be good at. The same is the case for the canoe club that we associate due to semantic knowledge with water that is contradictory to the very concept of a desert. It is the opposite of what we would expect. Irony is about **expectations**, which is what I will be concerned with in this chapter.

For situational irony, therefore, applies what was discussed for irony and humor likewise: We perceive it to be funny, when there is a direct contradiction in a history someone is telling us, where something apparently **incongruent** appears. Our expectations proceed language processing (top-down processing). What amuses us is when we detect that our expectations have led us on the wrong track. Which makes it necessary to revise our interpretation.

Jokes of this kind are like garden-path sentences, or as Zhao (2011, pg 179) calls it: "garden-path irony", which is another kind of logical mechanism Attardo and Raskin (1991) mention. The following is an instance of garden-path or false priming, to repeat their example that I mentioned already (Attardo and Raskin, 1991, pg 306):

(55) The astronomer married a star.

About their first parameter 'language' they wrote that the essential part of a joke either is based on ambiguity or contradiction (Attardo and Raskin, 1991, pg 299). And how this is told is the narrative strategy, where "missing links that the hearer must and can reconstruct" (Attardo and Raskin, 1991, pg 301) violate deliberately the Gricean maxim of quantity. By obviously not obeying the principle of quantity the utterance leaves gaps for the hearer to reconstruct. Just like for irony there is a missing link: the contradiction between what is said and what is meant, that can have the form of being the opposite of each other. The next parameter they propose is the 'target', which is analogous to the victim of sarcasm proposed by echoic theories. I argued that sarcasm is a special instance of irony where a critical attitude is conveyed but that for other uses of irony it does not have to be the case that there is a victim. It is very interesting that Attardo and Raskin note that the victim parameter is the only one that is optional. A distinction, they add, that has been made by many scholars on humor theory (Attardo and Raskin, 1991, pg 302). Together with 'situation' as fourth parameter and 'script opposition' as seventh, the sixth parameter 'logical mechanism' completes the seven parameters. By logical mechanism they mean the way to construct the contradiction, which would be for stereotypical cases of irony: opposition. So while "a joke must provide a logical or pseudological justification of the absurdity or irreality it postulates" (Attardo and Raskin, 1991, pg 307) irony must make its irreality in the existing situation transparent. So the script opposition as a "chunk of structured semantic information" (Attardo and Raskin, 1991, pg 307) is the interpretation evoked in the hearer. Only together with that a joke, as well as irony, is complete. This is also why a three-step grounding process makes so much sense for irony, because the speaker of the utterance must confirm whether he or she is all right with what his or her listener made with the utterance.

Computational irony What we saw so far, was that in psycholinguistic terms irony comprehension is guided by intentions that we attribute to the speaker. Our expectations guide this process to make it as efficient as it apparently is. This is incorporated in models on computational irony that

try to model irony detection in automatic language processing. Which is why I will review two ideas like that now.

Wallace et al. (2014) used an irony corpus consisting of comments from the social-news website "reddit" to investigate in the role of context. Students were asked to label utterances as ironic or not ironic. If they were not sure they could ask for further context, the context that was then provided consisted of previous comments of this user, the discussion that the user commented on, and other comments on the topic (Wallace et al., 2014, pg 514).

What they found was that for the comments that finally ended up classified as ironic all three different annotators more often asked for additional context than for the non-ironic comments. They conclude that the utterance itself and its intrinsic features are not enough to distinguish ironic from unironic comments (Wallace et al., 2014, pg 515). But if the annotators had no clue whether a comment was meant ironic or not would they have not needed to request context equally often for comments that finally were judged as non-ironic than ironic? Doesn't this show that annotators needed context to be sure if an utterance they already classified as potentially ironic was indeed ironic, rather than to differentiate between ironic and unironic? An example from their corpus: A user wrote to the issue of American politics on Republican senator Ted Cruz: "Great idea on the talkathon Cruz. Really made the republicans look like the sane ones" (Wallace et al., 2014, pg 514). This could either be the literally meant statement by a Ted Cruz supporter or an ironic comment by a Ted Cruz critic. And due to this ambiguity Wallace et al. (2014) conclude all three annotators indeed requested more context to decide. But there are clear hints in the utterance itself that this is more likely meant as ironic rather than literal comment: The word "talkathon" itself is negatively connotated as a word-play between *marathon* and *talk*, describing a unpleasantly long talk or discussion, this combined with the praise "great idea" produces a mismatch, that hints strongly to ironic use. In my opinion, what the more frequent request for contextual knowledge they found in their corpus study indicates is the ambiguous concept irony is and annotators asked for further support for an ironic interpretation they already had, due to the utterance itself.

Interestingly, in reviewing classical theories of irony, Wallace (2015) mentions the criticism of Sperber and Wilson at the account of Grice, that reads as follows: if one says *It's raining just a bit* not the negation as proposed by Grice is meant. Rather: "Of course, highlighting the absurdity of the notion that it is raining 'just a bit' also serves as an indirect way of communicating 'it is down-pouring'" (Wallace, 2015, pg 469). Wallace bases his proposal on the observation that the hearer, to successfully reconstruct the

ironic meaning, has to know something about the speaker. Here the pretense theory comes into play again, since it states two audiences: one that understands the irony, and another that just gets the literal meaning. Unless the speaker does not give clear hints, or the situation is obviously ironic is a subtle way of proposing an alternative meaning to those that have enough information to infer the opposite while others might just stick to the surface meaning (Wallace, 2015, pg 470). Therefore, Wallace proposes a computational model that includes the speakers environment to detect irony. What is meant by this is a hearer internal model of the speaker equipped with the values and attitudes of the speaker towards certain issues, combined in "aspects" (Wallace, 2015, pg 476). What Wallace sees as essential for modeling ironic speech correctly in a computational environment is context about the speaker and the situation to allow the hearer to construct aspects that reflect beliefs about these two variables. One could say these aspects are assumptions or prejudices on what is true about the speaker and the situation in the actual world that are good enough to detect a mismatch between our representations or aspects and the utterance (Wallace, 2015, pg 476). This speaker-model plus features regarding the literal proposition being expressed are assumed to work together. Previous models that were just concerned with the utterance-internal semantic features of an utterance came already to fairly good results in detecting irony, though, of around 70 percent precision (defined as "the fraction of sentences classified by the algorithm as ironic that in fact were") and accuracy (defined as "the total fraction of utterances correctly classified"). (The model by Davidov and Rappoport (2006) achieved for utterances from "amazon" 91 percent precision and 76 percent recall, and for a twitter corpus 72 percent precision and 44 percent recall, while the model from Tepperman et al. (2006) detected whether utterances saying "yeah right" were ironic or not with an 87 percent accuracy and F-measure of 70 percent (F-measure is the mean of precision and recall). (Wallace, 2015, pg 473).)

Wallace calls his pragmatic context model a "probabilistic user-model that captures expectations" (Wallace, 2015, pg 478). He means by this, for example, to consider the speaker's browser history to detect whether an utterance was meant ironic or not (Wallace, 2015, pg 476). (Someone normally searching for eco-friendly fair traded clothes writing "I think H&M is the best" might probably mean it ironically.) Or even, constructing a model of how the speaker normally writes and compare this language model of the speaker with the utterance in question to see whether a mismatch occurs (Wallace, 2015, pg 479). (Someone not using a lot of expressive language might be sarcastic when writing the answer "O.M.G.!!! This is exciting!" 'after a not so exciting fact.)

What I consider so interesting about this is the attempt to model expectation in a computational way and that these models combine the semantic information of the utterance itself with pragmatics to get the most precise irony detection.

Kao and Goodman (2015) use the same kind of logic. They also attempt to formalize how expectations guide our understanding of communication into a computational model about ironic speech. They understand “language understanding as recursive reasoning between speaker and listener” and understand the current speech act as question under discussion (QUD) that reminds us of various labeling models reported in the chapter about how theories tried to model common ground. Another parallel to the model by Wallace et al. (2014) is that they also incorporate a internal model of the speaker by the listener. They define two dimensions, arousal and valence in a two-dimensional affect space. They let the model judge statements like *The weather is amazing!* be judged with the help of a context picture about the weather and the emotion of a speaker, that was ranging on a Lickert scale from terrible, bad, neutral, good, to amazing. “We found a basic irony effect, where utterances whose polarities are inconsistent with the polarity of the weather context are rated as significantly more ironic than utterances whose polarities are consistent with the weather context.” And: “This suggests that participants lay judgments of irony align with its basic definition: utterances whose apparent meanings are opposite in polarity to the speakers intended meaning.”

Their model’s predictions of the weather state judged by the utterance matched humans’ interpretation by 86 Percent. After comparing these results with simpler models, that for example did not incorporate the emotional states, the complete model outperformed those. What it took into account, was: prior knowledge, literal meaning of the utterance, and a two-dimensional affect space, with valence and arousal. They conclude that like this pragmatic reasoning can successfully be modeled computationally. Which means that this reasoning for them uses shared background knowledge and the speaker’s affective goals.

Chapter 5

General discussion

I would like to discuss three hypotheses that are central to my thesis.

- Irony is a two-step process, where meaning is created in a reciprocal process between the speaker and the listener.
- Irony communicates meaning in that it provokes an absurd context that clashes with the actual context shared by the interlocutors.
- Why the meaning of irony is so difficult to pin down is because not only is it pragmatic like implicatures or presuppositions are, but it is also dependent on the ability and willingness of speaker and hearer to add this layer of meaning, which depends on their relationship and ease with which they communicate, but also to the various clues present in the dialog.

Talking about the theory, we discussed the communicative function of irony and that there are mainly two viewpoints to it: The theories following Grice (1967) in understanding irony as a figure of speech where the opposite of what is said is meant. On the other hand, there are the theories that see irony as mention or pretense, where always an attitude towards the ridiculed issue is conveyed. Representing these accounts is the relevance theory, with which Sperber and Wilson (1981) argued against Grice that there are other instances of irony where something different but not the opposite of what is said is meant. The graded salience hypothesis, as Neo-Gricean account, assumes that irony is indirect negation, functioning, as the retention hypothesis describes it, by comparison of the literal and non-literal meaning and pointing attention to the difference between the two. Pointing in the same direction, the relevant inappropriateness hypothesis by Attardo (2000)

understands irony as purely pragmatic phenomenon that always has to be derived.

To do so the interlocutors have to share a certain situation. Just like a joke that often needs a certain common cultural background, irony needs the interlocutors to have a situational background. Therefore, I reviewed an account on humor to show that some jokes and irony are based on the same structure: Opposition, derived from contradiction, as opposition is a special case of contradiction. For example, West (2013) examined the irony of Stephen Colbert an American comedian that played for years the role of an ultra-conservative supporter of the Republican party. He ridiculed Republicans by exaggerating their world view and by exposing himself to ridicule as playing the most burning Republican, as a result he turned its meaning to the opposite and mocked their ideas themselves:

Colbert's oral testimony is replete with his brand of irony. He refers to his 'vast experience spending 1 day as a migrant farm worker' and when speaking of our dependence on those workers, proposes that 'the obvious answer is for all of us to stop eating fruits and vegetables'. Ultimately turning to comic hyperbole, he laments having to spend the day bending over to pick beans: 'It turns out – and I did not know this – most soil is at ground level. If we can put a man on the moon, why can't we make the earth waist high?' (West, 2013, pg 16)

What this example shows so nicely, is that one cannot confuse the means to produce irony with the effect this has. Which is just the relationship we found between irony and implicature. Different tropes are used here, as West (2013) points out, is hyperbole here that makes the literal meaning implausible and producing irony that has the effect of inverting the utterance so that what he finally says is the opposite. In the end, by pretending to be a conservative he makes their ideas look ridiculous and actually argues for the exact opposite viewpoint. What he communicates is the opposite of what he says: He has no experience whatsoever about what it means to be a farm worker, after working as such for one day. Also, their job is an immensely important one because everyone has to eat fruit and vegetables. Finally, even a nation that can put a man on the moon should praise their jobs and appreciate how hard it is since no one can put the earth waist high.

To sum up the theory, I would support the standard pragmatic model, and in line with Grice, who said irony is the negation of the literal meaning, the relevant inappropriateness hypothesis can explain some features of irony quite convincingly, for example, why sarcasm is much more often used than kind irony. In my opinion, the graded salience hypothesis grasps why we use

irony: to communicate the present contradiction that informs and amuses us. The graded salience hypothesis also assumes irony to be a case of indirect negation. Another point I tried to make is in line with the pretense theory that postulates two kinds of audiences: one that detects the irony and one that understands only the literal meaning. I assume that irony – in its more subtle cases – functions like a proposed additional layer of meaning for those that have enough information and are willing to detect it.

Wallace (2015, pg 480) calls that "post-modern modes of irony" in contrast to stable irony that are cases where the speaker avoids to position him- or herself in any way. Then it just blurs any meaning and makes it dubious. Which brings us to the psycho- and neurolinguistic evidence.

We saw at the end of the first chapter in (2.6.3) where studies and theories overlap. For Grice the "said and the unsaid" (Chierchia 102) are important and both contribute to meaning because reasoning about the choices the speaker made allows the hearer to reconstruct the communicative intention of the speaker. Which was exactly what was discussed in the chapter on studies among ASD participants. This is the step that has been found to be difficult for them and might be where they lack to construct inferences about the speakers intention. To do this, a presupposed shared background has to be present for the speaker and the hearer to be 'on the same page' about what their shared believes are. I argue that the most distinct cases of irony function due to a structural contradiction in the utterance itself, or between the utterance and the context, or between the utterance and what the speaker assumes to be believed by the speaker, that forces us to reconsider and construct a ridiculous background. This structure parallels jokes. We have seen that many jokes function with a figure-ground reversal. Like the joke about ten people that turn a light bulb in because they turn the table the person putting the light bulb in is standing on, instead of turning the light bulb. This is the same absurdity than bringing the ground above ground level, as Colbert proposed.

If irony is just stating the opposite of its literal meaning, why don't we just say the opposite then? Because irony allows the speaker to make the contradiction obvious, which communicates something itself, namely the clash between the real world and the one we were forced to reconstruct to make sense of the ironic utterance. It is the thinking outside the box, that reminds us of the box that we are normally thinking in, that might be us making stereotypical and racist jokes about the stupidity of polish people (as the example of a joke from Attardo and Raskin about the light bulb). In other words: what irony communicates is the alternative background it evokes and the mismatch between it and the present real background.

Another example comes from Zhao (2011, pg 177) she cites Austen (1996, pg 3): Mrs Bennet: " 'Don't keep coughing so, Kitty, for heaven's sake! Have a little compassion on my nerves. You tear them into pieces.' After Mrs Bennet's mood lightens Mr Bennet says: 'Now, Kitty, you may cough as much as you choose.' " Zhao analyzes Mr Bennet's remark as an echo to the previous utterance of his wife, Mrs Bennet, which is in line with echoic theories. But this has not to be seen as an echo to understand the irony. Since common knowledge suggests that coughing is nothing that you can choose to do, this contradiction that points to an ironic meaning forces one to think about a background where coughing can be chosen, just like Mrs Bennet implied.

In terms of how implicature and irony interact, implicatures can be used to derive irony, but it is one possibility among others, that are for example exaggeration, understatement, or omission of relevant information. Irony and jokes work with inconsistencies in a certain context. Which is why I tried to show the similarities between jokes and irony, but one has also to pin down what distinguishes them: Both amuse us because they surprise us. The element of surprise, however, typically differs: Jokes show us the truth in something obviously ridiculous while irony shows the untruth about something that could just as well have been true.

Jokes, irony and implicatures enrich meaning because the hearer is forced to construct a fitting context against which the ironic utterance makes sense. Just like the punch line of a joke forces us to construct an absurd background that is consistent with the joke and that makes us laugh. But while (scalar) implicatures are a strengthening device irony could be a weakening device. Chierchia (2006, pg 3, 17) provides a framework that shows that some implicatures lead to strengthening of meaning but not all of them, some also lead to weakening in that they could widen a domain in which to search for an alternative, which leads to the weakening of the meaning of a statement (Chierchia, 2013, pg 25). This is similar to the "tinge hypothesis" that Filik et al. (2015) discuss in their study. They found evidence for the claim that sarcasm (as which they understand ironic praise and ironic criticism) weakens praise and criticism, because a literally negative utterance meant ironic is "tingued" by the literally negative meaning, while a literally positive utterance meant as sarcasm is "tingued" by the literally positive meaning. So sarcastic praise was perceived as less positive than literal praise while sarcastic criticism was perceived less negative than literal criticism. Another point that psycho- and neurolinguistic studies provide evidence for is that for irony both, the literal and non-literal interpretation remain salient to construct the mismatch between the two. This is in line with proposals such as the retention hypothesis that assume a two-step process for irony interpretation, since

only if the literal and opposite, non-literal interpretation are accessible the difference between them can become obvious. Furthermore, irony has to be the non-salient interpretation always, since it would not be surprising and humorous if it were salient.

My assumption about irony follows Grice's account in that it is understood by the hearer that the speaker obviously flouts one maxim but still obeys the cooperative principle.

Since for Grice irony only involves conversational implicatures and they are derived by flouting the maxim of quality I discussed to which extend implicature and presupposition show similarities to irony.

Even though I argued against echoic theories, I assume that irony can come at cost of a "victim" but only in the case of sarcasm. That irony and sarcasm are not distinct cases but different shades of the very same phenomenon was supported by psycholinguistic studies on how children comprehend irony.

What I also tried to show in the chapter on psycholinguistic research is that my claim about the ironic meaning being opposed to the literal meaning is what the retention hypothesis predicts for non-salient irony. This is supported by psycholinguistic research, where it has been shown that both interpretations remain active. Also evidence points to the fact that understanding irony seems to be about putting many different clues together and combining them to a strong hypothesis that something different is meant than the literal meaning of the utterance. This might be something that is difficult to impaired patients due to restricted memory resources in general. For ASD participants the incorporation of different clues to construct a consistent utterance relative to a relevant background might be too much, because pointed right to a sarcastic or ironic meaning they indeed are able to understand irony. It was also clearly shown that second-order theory of mind, to attribute mental states to one's opposite and draw inferences from that about what the speaker wants to communicate, is necessary to understand irony. Interestingly RBD patients and ASD participants did not make arbitrary mistakes when the task was to choose the correct punchline or to decide whether someone was lying, or being ironic. They were shown to fail at choosing the option that would become consistent if one would alter the context accordingly. Which means, jokes are inherently consistent and irony makes sense if you construct a fitting context. To alter the actual context to a context that would incorporate the punchline or the irony seems to require a second-order theory of mind. Since that is needed to interpret a joke or an ironic utterance in so far as it attributes to the speaker that he or she had some cooperative intention in mind and wasn't plainly lying. Which is a process guided by one's own expectations as well.

That is why we then saw an idea how pragmatically enriched meaning can be modeled computationally. To me these models are interesting because they show how context, a representation of the speaker by the hearer, and expectations interfere in the comprehension of an ironic utterance all at once. These expectations guide to what we, at best, are able to draw from all these knowledge resources: to make sense out of a seemingly incongruent statement.

Chapter 6

Conclusion

This thesis does not provide an answer to the question, what the meaning of irony is. But, from a theoretical point of view, theories on irony have been sorted and discussed so they can provide new insight. Also, the discussion of psycho- and neurolinguistic evidence, as contradictory as it sometimes is, has been reviewed again in order to find consistent findings and views on irony. So I can provide some answers or ideas to the following questions:

Why is sarcasm so much more often than kind irony? The most intriguing explanation for this very much obvious, but still puzzling feature of irony comes from Attardo (2000, pg 796), who assumed that irony is as Zhao (2011, pg 180) said “in the eye of the beholder”. So irony not only always has to be inferred but it is also a choice of the hearer. Therefore, since as speaker I do not know with certainty whether my hearer will get my irony, positive irony, which is saying something negative meaning something positive is more dangerous a concept than sarcasm, which involves saying something positive meaning something negative. If my opposite does not construct the ironic meaning and my intended insincere utterance fails, the hearer stops at the positive meaning, which causes no harm.

That points to the fact that in this thesis I argued for a two-stage processing of irony. This follows the retention hypothesis, which states that irony is constructed out of the literal plus the non-literal meaning.

At this point we also saw that irony can be deleted just like conversational scalar implicatures. Implicatures and irony can be the same, they don't have to be though. This is where Grice's theory cuts short. Irony and implicature stand in a relationship with each other: implicatures can be used to be ironic but also other means like understatement, exaggeration or the omission of relevant information can derive irony. We saw that, just like for jokes and implicatures, for irony the hearer needs to enrich meaning with the knowledge resources he or she has to derive a meaningful description of a situation.

So, does irony enrich meaning, just like implicatures do? This idea is intriguing even though we cannot explain irony like this. This is due to several factors, for one, so far, there is no theory of pragmatic scales, which would be needed to discuss irony as exhaustification or functioning somehow similar to that. Furthermore, just like implicatures do not all lead to strengthening via enrichment with relevant alternatives, irony does not always apply the same mechanisms. But as its effect it evokes a certain background that needs to be constructed to implement the ironic utterance. Sometimes it forces the hearer to construct a figure ground reversal, and sometimes it strengthens meaning as it points to the relevant alternatives evoked from discourse.

What is presupposed between the interlocutors to communicate an ironic meaning successfully? For irony to work, no accommodation process as for presupposition is possible. For the interlocutors to understand one another, what is shared by the interlocutors, whether you want to call it common ground, background, or context, has to be set. Since it works as the benchmark to construct an alternative background against which the ironic remark makes sense.

And do some jokes and irony employ the same logical mechanisms? As we have seen from psycho- and neurolinguistic evidence, compared to lies, the difference between an arbitrary surprise and a “real” joke ending cannot be completely arbitrary. So I would assume that in the most cases irony and jokes employ the very same logical mechanisms. Which would make it fruitful for future research to search for further parallels between jokes and irony in processing.

The argumentation of this thesis can be seen to belong to the current turn in pragmatics where increasing evidence has been found in support of a view that understands pragmatic inferences to intrude in semantics.

Bibliography

- Attardo, S. (1989). A multiple-level analysis of jokes. *Contributed section in Newsletter, HUMOR: International Journal of Humor Research*, 2:438–439.
- Attardo, S. (2000). Irony as relevant inappropriateness. *Journal of pragmatics*, 32(6):793–826.
- Attardo, S. and Raskin, V. (1991). Script theory revis (it) ed: Joke similarity and joke representation model. *Humor-International Journal of Humor Research*, 4(3-4):293–348.
- Austen, J. (1996). *Pride and prejudice*. Penguin Group London.
- Austin, J. L. (1962). *How to do things with words*. Oxford University Press.
- Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a theory of mind? *Cognition*, 21(1):37–46.
- Bavelas, J. B., De Jong, P., Korman, H., and Jordan, S. S. (2012). Beyond back-channels: A three-step model of grounding in face-to-face dialogue. In *Proceedings of Interdisciplinary Workshop on Feedback Behaviors in Dialog*. Stevenson, Washington, pages 5–6.
- Beyssade, C. and Marandin, J.-M. (2006). The speech act assignment problem revisited: Disentangling speakers commitment from speakers call on addressee. *Empirical issues in syntax and semantics*, 6:37–68.
- Brône, G. (2012). Humour and irony in cognitive pragmatics. *Cognitive Pragmatics*, 4:463.
- Cheang, H. S. and Pell, M. D. (2006). A study of humour and communicative intention following right hemisphere stroke. *Clinical linguistics & phonetics*, 20(6):447–462.

- Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the logicity of language. *Linguistic inquiry*, 37(4):535–590.
- Chierchia, G. (2013). *Logic in grammar: Polarity, free choice, and intervention*, volume 2. Oxford University Press.
- Clark, H. H. and Brennan, S. E. (1991). Grounding in communication. *Perspectives on socially shared cognition*, 13(1991):127–149.
- Clark, H. H. and Gerrig, R. J. (1984). On the pretense theory of irony. *Journal of Experimental Psychology: General*, 113(1):121–126.
- Creusere, M. A. (2000). A developmental test of theoretical perspectives on the understanding of verbal irony: Children’s recognition of allusion and pragmatic insincerity. *Metaphor and Symbol*, 15(1-2):29–45.
- Davidov, D. and Rappoport, A. (2006). Efficient unsupervised discovery of word categories using symmetric patterns and high frequency words. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. Sydney*, pages 297–304. Association for Computational Linguistics.
- Eggers, D. (2000). *A Heartbreaking Work of Staggering Genius*. Simon & Schuster New York.
- Filik, R. and Moxey, L. M. (2010). The on-line processing of written irony. *Cognition*, 116(3):421–436.
- Filik, R., Turcan, A., Thompson, D., Harvey, N., Davies, H., and Turner, A. (2015). Sarcasm and emoticons: Comprehension and emotional impact. *The Quarterly Journal of Experimental Psychology*, pages 1–17.
- Frege, G. (1892). Über sinn und bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100(I):25–50.
- Frege, G. (1948). Sense and reference. *The philosophical review*, 57(3):209–230.
- Frith, U. and Happé, F. (1994). Language and communication in autistic disorders. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 346(1315):97–104.
- Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and symbol*, 15(1-2):5–27.

- Giora, R., Fein, O., and Schwartz, T. (1998). Irony: Graded salience and indirect negation. *Metaphor and Symbol*, 13:83–102.
- Giora, R., Zaidel, E., Soroker, N., Batori, G., and Kasher, A. (2000). Differential effects of right-and left-hemisphere damage on understanding sarcasm and metaphor. *Metaphor and Symbol*, 15(1-2):63–83.
- Grice, H. P. (1967). Logic and conversation. *Syntax and semantics*, 3:41–58.
- Grice, H. P. (1978). Some further notes on logic and conversation. *Syntax and Semantics: Pragmatics*, 9:113–127.
- Happé, F. G. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, 48(2):101–119.
- Heim, I. (1990). Presupposition projection. In *Reader for the Nijmegen Workshop on Presupposition, Lexical Meaning, and Discourse Processes. University of Nijmegen*, volume 31.
- Kao, J. and Goodman, N. (2015). Lets talk (ironically) about the weather: Modeling verbal irony. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society. Pasadena, California*, pages 1051–1056. Cognitive Science Society.
- Lucas, G. (1977). Star wars: Episode iv - a new hope. revisited fourth draft.
- Malamud, S. A. and Stephenson, T. (2015). Three ways to avoid commitments: Declarative force modifiers in the conversational scoreboard. *Journal of Semantics*, 32(2):275–311.
- McDonald, S. (2000). Neuropsychological studies of sarcasm. *Metaphor and Symbol*, 15(1-2):85–98.
- McDuffie, A., Yoder, P., and Stone, W. (2006). Fast-mapping in young children with autism spectrum disorders. *First Language*, 26(4):421–438.
- Monetta, L., Grindrod, C. M., and Pell, M. D. (2009). Irony comprehension and theory of mind deficits in patients with parkinson’s disease. *Cortex*, 45(8):972–981.
- Nakassis, C. and Snedeker, J. (2002). Beyond sarcasm: Intonation and context as relational cues in childrens recognition of irony. In *Proceedings of the Twenty-Sixth Boston University Conference on Language Development. Cascadilla Press, Somerville, Massachusetts*, pages 429–440.

- Panizza, D., Chierchia, G., and Clifton, C. (2009). On the role of entailment patterns and scalar implicatures in the processing of numerals. *Journal of Memory and Language*, 61(4):503–518.
- Potts, C. (2005). *The logic of conventional implicatures*, volume 7. Oxford University Press.
- Raskin, V. (1985). *Semantic mechanisms of humor*.
- Reimer, M. (2013). Grice on irony and metaphor: Discredited by the experimental evidence? *International Review of Pragmatics*, 5(1):1–33.
- Russell, B. (1905). On denoting. *Mind*, 14(56):479–493.
- Sabbagh, M. A. (1999). Communicative intentions and language: Evidence from right-hemisphere damage and autism. *Brain and language*, 70(1):29–69.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*, volume 626. Cambridge University Press.
- Searle, J. R. (1975). *Indirect speech acts*. Academic Press New York.
- Sperber, D. and Wilson, D. (1981). Irony and the use-mention distinction. *Philosophy*, 3:143–184.
- Sperber, D. and Wilson, D. (1995, 2nd edition). *Relevance: Communication and cognition*. Wiley-Blackwell Oxford.
- Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic*, 2(4):447–457.
- Tepperman, J., Traum, D. R., and Narayanan, S. (2006). ” yeah right”: sarcasm recognition for spoken dialogue systems. In *INTERSPEECH. International Conference on Spoken Language Processing*. Pittsburgh, Pennsylvania.
- Van der Sandt, R. A. (1992). Presupposition projection as anaphora resolution. *Journal of semantics*, 9(4):333–377.
- Wallace, B. C. (2015). Computational irony: A survey and new perspectives. *Artificial Intelligence Review*, 43(4):467–483.

- Wallace, B. C., Do Kook Choe, L. K., Kertz, L., and Charniak, E. (2014). Humans require context to infer ironic intent (so computers probably do, too). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, Maryland*, pages 512–516.
- Wang, A. T., Lee, S. S., Sigman, M., and Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: the role of prosody and context. *Brain*, 129(4):932–943.
- West, T. L. (2013). Meta-indexing the abstraction of irony and the structural differential. *et Cetera*, 70(1):13.
- Winner, E., Brownell, H., Happé, F., Blum, A., and Pincus, D. (1998). Distinguishing lies from jokes: Theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain and language*, 62(1):89–106.
- Zhao, H. (2011). A relevance-theoretic approach to verbal irony: A case study of ironic utterances in pride and prejudice. *Journal of Pragmatics*, 43(1):175–182.
- Zimmermann, T. E. and Sternefeld, W. (2013). *Introduction to semantics: an essential guide to the composition of meaning*. Walter de Gruyter Berlin/New York.

Zusammenfassung

Diese Masterarbeit hat zum Ziel, Ironie von einem kognitionswissenschaftlichen Standpunkt aus zu beschreiben. Klassische und neuere linguistische Theorien zu übertragener Bedeutung und Humor werden beschrieben und verglichen. In der Pragmatik wird Ironie von drei verschiedenen Standpunkten aus gesehen: Grice definierte Ironie als Implikatur, Ironie kann als mitgemeinte Bedeutung semantisch als das Gegenteil des Gesagten betrachtet werden. Dem entgegen, stehen Ansätze, die Ironie rein pragmatisch betrachten, dessen Bedeutung entstehe erst durch die Interpretation des Gesagten auf der Suche nach der relevantesten Information. Neuere Ansätze greifen Grice auf, insofern dass Ironie indirekte Verneinung ist und als eigentlich unangemessene Bedeutung so besondere Aufmerksamkeit auf sich zieht. Wo genau überlappen sich Ironie und Implikatur und wo unterscheiden sie sich? Außerdem wird Ironie in der Folge verglichen mit Präsuppositionen, also dem Vorausgesetzten an Information, und der Theorie, dass Gesprächspartner einen geteilten Informationshintergrund, einen Common Ground, haben müssen.

Die eingangs vorgestellten Theorien werden dann in Bezug auf aktuelle psycho- und neurolinguistische Studien diskutiert. Dort findet sich Evidenz für Ironie als zweistufigen Prozess, der es erfordert verschiedene Hinweise zu verarbeiten und dessen Verstehen ein sogenanntes Konzept der Theory of Mind erfordert. Außerdem scheint entscheidend, dass vom Gesprächspartner die kommunikative Absicht des Sprechers erkannt wird. Dies hängt sowohl mit einem Common Ground, also dem inhaltlich eingegrenzten Feld des Gesprächs, zusammen, als auch mit Grices Idee eines Kooperationsprinzips, auf das sich beide Gesprächspartner einigen. Neben der Situation und der Beziehung zwischen den Sprechenden, spielt beides eine Rolle beim Verstehen übertragener Bedeutung.

Letztendlich werden mögliche Erkenntnisse zum Ironieverstehen diskutiert, die das Feld der experimentellen Pragmatik betreffen, wo weitere Forschungsfragen entstehen. Die experimentelle Pragmatik untersucht mitgemeinte Bedeutung, wie Implikaturen, als Anreicherung von Bedeutung. Zwischen den Gesprächspartnern wird verhandelt, was gemeint ist, indem sowohl das Gesagte, als auch das nicht-Gesagte einbezogen und mit Logik gegeneinander abgewogen werden. So erschließt sich die gemeinte ironische Bedeutung dem Hörer, woraufhin dieser oder diese dem Sprecher signalisiert, dass er oder sie verstanden hat und möglicherweise erneut vom Sprecher bestätigt wird, ob er oder sie richtig verstanden wurde. Es findet ein Abgleich zwischen Gemeintem und Verstandenem statt, bis beide möglichst deckungsgleich übereinander liegen.