Use of repair on Reddit. Providing an overview of the use of repairs by the users of Reddit.
Abstract

1 Introduction
1.1 Reddit as research subject
1.2 Research questions
1.3 Hypotheses
1.4 Reading guide

2 Theoretical framework
2.1 Social Media and Reddit as a social medium
2.2 Describing Reddit
   2.2.1 Front page and subreddits
   2.2.2 Determining the order of the posts
   2.2.3 Upvoting and downvoting
   2.2.4 Rules and reddiquette
   2.2.5 Editing a post on Reddit
2.3 What is a repair?
   2.3.1 Repairs
   2.3.2 Creating and monitoring of repairs
   2.3.3 Types and places of repair
   2.3.4 Covert repairs
   2.3.5 Overt repairs
   2.3.6 Self-initiated and other-initiated repairs
   2.3.7 Repair after next turn
   2.3.8 Private versus public editing in CMC/Reddit
   2.3.9 Other uses of repairs
   2.3.10 Distinguishing between an edit and repair on Reddit
2.4 From the collaborative model of communication to grounding in CMC
   2.4.1 Grices maxims
   2.4.2 From the cooperative principle to the collaborative model
   2.4.3 From collaboration to grounding the conversation
   2.4.4 Media richness theory
   2.4.5 Dealing with the limitations of CMC
   2.4.6 Repair in chat room interaction
   2.4.7 Accountability
2.5 Difficulties of using Reddit as a user and researcher
2.5.1 Differences in repair use in Reddit and spoken interaction 42
2.5.2 Repairs and edits in CMC and Reddit 43
2.6 Additions to the classification of Levelt 46

3 Research method 54
3.1 Mining Reddit 54
  3.1.1 Using the API 54
  3.1.2 Choosing the subreddits 56
3.2 Extracting the repairs from the raw data 59

4 Results 60
  4.01 Examples of repairs 61

5 Conclusion 66
  5.0.1 Answers to the research questions 66
  5.0.2 Summary 72
5.1 Discussion and suggestions for further research 73
  5.1.1 Critical remarks regarding the research 73
  5.1.2 Suggestions for further research 74

Literature 77
Abstract

Conversation is never without error. Repairing these errors is done in all forms of communication, and in a variety of ways. We have mined data from Reddit and researched the use of repairs in the conversations on Reddit. Reddit, being a middle ground between a strict online forum and an online chat program, has features of both. Rather than an in depth research of repairs, we aim to provide an overview of repair use on Reddit. We mined data from five popular subreddits and categorized them according to Levelt's categories and added our own. The results show that users of Reddit use repairs in an unique way: not only for error correction in the broadest sense of the word, but also as a tool to improve an ongoing conversation. This is done, for example, by providing sources for claims amongst others. This suggests a departure from the notion that a repair is an unwelcome interruption of a conversation, a notion widely held in face-to-face communication, at least in the case of Reddit.
1 Introduction

Communication is something we do every day in several different ways without giving it much thought. It seems to come so naturally that researching it could appear to be simply pushing at an open door. However, communication warrants in depth research, especially when processes seems so blatantly obvious. Take the following: communicating is rarely done without error, from small errors in words to completely misunderstanding or misconstruing the intended message. But try to count how many little errors you make in your day-to-day conversation and you will most likely be amazed by the sheer number. What's even more amazing is the fact that despite the high number of errors you will find the number of times you were unable to have a proper conversation was much lower, or perhaps non-existent. Perhaps you communicated successfully because your partner correctly guessed what you were trying to say or perhaps you either discovered or had your attention brought to an error and repaired it. Those repairs are the subject of this thesis. We will look at these repairs from a collaborative viewpoint, rather than taking a more technical approach focused on the literal transferring of the message. The collaborative model sees communication not as an exchange of individual utterances, but as an ongoing process in which both speaker and listener are equally important to the success of the conversation. Other models such as the famous 1949 model of Shannon and Weaver, commonly known as the sender-message-receiver model, are more focused on the technical aspect, rather than on the content and understanding of the message. Obviously communication can be analysed from more than just a technical point of view. So rather than analysing the how of sending/receiving the message, we focus on the understanding, or misunderstanding of the message sent. Specifically, we look at the way that users use the edit function to repair their posts on the Social Network-site Reddit.com. By mining data from Reddit we were able to gather a corpus of written text and we analysed the ways in which users edited their own posts and the reasons they gave for it. To categorise the data we have gathered we turned to the classification of Levelt (1983) and added our own categories based on an observation of Reddit spanning 2014 to 2015.
1.1 Reddit as a research subject

Why choose Reddit as a research subject, why not for example Twitter? A simple Google Scholar query suggests that Twitter is a far more popular social medium to research. Google Scholar returns approximately 3.32 million results for the term “twitter” (minus the sentences “share on twitter” and “add to twitter”) versus approximately 245,000 results for the term “reddit” (minus the sentences “add to reddit” and “share on reddit”). Additionally; the results for Twitter provide more relevant links to research done on Twitter than for Reddit, where the Journal of Bone & Joint Surgery appears twice on the first page of results. Research regarding the use of repairs on Reddit wasn’t found.

Of course, a prominent reason for research on Reddit could then be that not much research has been done on Reddit, and that this somewhat dark territory is worth exploring simply because no one has done so yet. An underlying reason, certainly, but not the primary one. Reddit was primarily chosen because of its characteristics and features that are close to unique for computer mediated communication (CMC). Combining elements of online chat and forums, Reddit is a social medium that provides an insight in CMC found in neither chat nor forums. For example; the tree-structure of the posts and replies visualise discussions and interaction not found in online forums or chat rooms where posts are grouped chronologically. Reddit can have a high speed of conversation comparable to online chat, at least in the more popular subreddits, but can also contain large amounts of text in one reply or posts posted more slowly as found in online forums.

Added to that is the relatively easy way large amounts of data can be gathered due to the popularity of the site. Reddit has a lot of traffic and content. According to ranking site Alexa, Reddit is the 33rd most popular site in the world and ranked higher than Pinterest.com and even Microsoft.com. Reddit even ranks 10th in the United States (Alexa, 2015). Reddit isn’t only popular, it also covers a diversity of subjects ranging from news to sports such as fencing. Data, effectively the posts made by users of Reddit, so called redditors is easily mined via an API (Application Programming Interface) and the complete code of Reddit is open-source.

The combination of these factors, the uniqueness of communication on Reddit, the popularity ensuring a large amount of data, and the fact that this data is easily accessible and covers a broad variety of subjects makes Reddit a prime candidate for research, especially were the research covers conversation in the broadest sense of the word as Reddit isn’t made up of a collection of single posts and the occasional interaction (like Twitter) but of a large array of conversations. By researching Reddit’s use of repairs we will be able to provide insight in the way
redditors communicate with each other and get a better understanding of the way CMC is used by this large and diverse group of people.

1.2 Research questions

Miscommunication is a given thing in communication. Only if we are able to read each others minds we can be fully certain of each others intentions. When miscommunication occurs in a conversation the participants have to undertake some action in order to ensure successful communication, in other words; they have to repair their conversation. In 1983 Levelt conducted a study on repairs that served as a foundation for a number of other studies on the same subject. In short, Levelt researched the way people use repairs and categorized these. Research on the subject of repairs isn’t done solely by Levelt, influential other researchers such as Schegloff have also researched repairs, yet aren’t as suitable as a starting point for this research, this is explained in more detail in 2.6

Repairs aren’t unique to face-to-face conversation but are also found in online Computer Mediated Communication (CMC) as chat and online forums. Reddit combines features of both chat and forums and provides redditors with the option to repair their posts. Research done on Reddit is relative scarce and we were unable to find research regarding the use of repairs on Reddit. Being unable to find any research combining repairs in communication and Reddit, it is likely best to establish an overview of repair use on Reddit. By categorising a corpus of repairs from Reddit we will be able to answer the following research question:

Are repairs used on Reddit in a manner unique to Reddit or do they adhere to the conventions as seen in either CMC or classic papers on repairs?

To answer this main research question, we must find answers to the following sub-questions

1. Do the repairs in conversation in Reddit adhere to the classification and distribution of numbers of repairs found by Shegloff (1992,2014), Levinson (1983) and Levelt (1983)?

2. What are the conventions for repair in conversation and text-based CMC and does Reddit follow these conventions?
3. If Reddit doesn’t adhere to 1 or 2, how and why have the users of Reddit adapted their use of repair to the specific communication options of Reddit?

1.3 Hypotheses

Although Reddit is largely unresearched, the theoretical background behind the collaborative model was developed over several decades of research. Rather than starting from a completely blank slate, we can hypothesise about the results of our findings and the theoretical reasons behind them.

Levelt (1983) was one of the first to categorise repairs. Establishing a starting point by looking at the categorisation of repairs done by Levelt, we can answer how and if the repairs in conversation in Reddit adhere to his classification and quantification of repairs.

We hypothesise the following: it is probably unlikely that Reddit follows the categorisation of Levelt, or adheres to the distribution of the categories found by Levelt. The reason for that is twofold. First there is a difference between the face-to-face communication Levelt based his work on and the text-based, computer mediated communication that defines Reddit. Second, Reddit allows what is essentially a self initiated self repair after the post is placed, and part of the conversation, something impossible in talk, or even in chat. Assuming the repairs do not adhere to Levelt (1983) we hypothesise that users of Reddit developed their own unique reasons for repairing a post.

Finally, we also hypothesise that Reddit follows the conventions for repair in text-based CMC and that their use of repairs is adapted to the constraints and affordances of CMC and Reddit in particular. We can use the fact that Reddit gives us the meta data about the post being edited or not, combined with a user explaining his reason for an edit to analyse the edited posts to answer these questions.

1.4 Reading guide

This thesis is build up in five chapters. In the second chapter we will lay out the theoretical framework by describing Reddit, followed by the communicative theories used for the research. After that we will describe the difficulties concerning the use of Reddit from both a user and researcher's perspective. In chapter three we will set out the research method and cover the mining of Reddit and the extraction of repairs from the data gathered by the mining. Chapter four covers
the results found. Finally, in chapter five we will answer our research questions followed by a discussion and suggestions for further research.

2 Theoretical Framework

In this part we will lay out the theoretical framework of this thesis and explain amongst others the concepts of repairs and the collaborative model. In paragraph 2.1 we will cover the definition of a social medium is and why Reddit is one, even though it doesn’t adhere to one of the most prominent definitions of social media. This as Reddit can be seen as solely a link aggregator rather than a social medium to converse on. This will be followed by a description of Reddit in 2.2. Paragraph 2.3 covers what a repair is, and follows with a description of the types and places of repairs. Then we will cover Grice’s maxims of communication and his introduction of the cooperative principle, arguing that redditors could aim to adhere to Grice’s maxims by breaking them before moving on to Clark and Wilkes-Gibbs collaborative model, the communicative model in which we’ve placed this research. Lastly we will describe the concept of grounding the conversation and the difficulties associated with this via the introduction of the media richness theory and problems with interactional coherence in CMC. In this way we provide a funnel approach: from the abstract concept of the collaborative model to the concrete example of CMC on Reddit.

2.1 Social Media and Reddit as a Social Medium

Although popular in the USA according to its Alexa ranking, Reddit isn’t all that well known in the Netherlands and explaining exactly what Reddit is is often difficult. According to Wikipedia (2014) Reddit “is an entertainment, social networking service and news website where registered community members can submit content, such as text posts or direct links. Only registered users can then vote submissions "up" or "down" to organize the posts and determine their position on the site’s pages.”.

On its frequently asked questions page Reddit defines itself as “Reddit is a source for what's new and popular on the web. Users like you provide all of the content and decide, through voting, what’s good and what’s junk. Links that receive community approval bubble up towards #1, so the front page is constantly in motion and (hopefully) filled with fresh, interesting links.” (Reddit d, 2015)

Reddit is thus defined by Wikipedia as a Social Networking Service and by itself as a source
of user provided content. In this thesis we look at Reddit as a social medium and study the interaction between it's members, rather than analysing the content (links) provided by its members. Social Networking Service is a term often used interchangeably with Social Media. As a social medium, Reddit is not only host to a large amount of discussions about a variety of subjects by a global public but can also be analysed as a form of social interaction. Singer, Flock, Meinhart et al (2014) have paved the way in this aspect. Their research, conducted over a period of five years, shows that Reddit changed from a mere link providing site to an intricate social community. Before going into the specifics of Reddit, it is therefore wise to recapture on what Social Media is. Social Media is one of today’s buzzwords of the internet, and as a term came into being around 2005 (Lietsala, Sirkkunen 2008). Coming up with examples of social media and their use is rather easy: even more than in 2008, Social Media is more integrated into our everyday lives. Facebook, Twitter and many more examples are widespread and used in various ways. Social Media are used from keeping in touch with friends, family and colleagues to being used for polling or sound bites in the evening news. Defining what Social Media is, or what the characteristics of a Social Medium are more difficult. In their 2010 article Users of the world, unite! The challenges and opportunities of Social Media, Kaplan and Haenlein defined Social Media as:

“Social Media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content”.

This definition relies on the concepts of Web 2.0, the shift in thinking about the internet as a static set of pages that once made are only maintained by their original creator but rather by as a collection of continuously updating and changing websites developed by the users. Wikis such as Wikipedia are a prime example of the Web 2.0 way of thinking. There were online encyclopedias before Wikipedia, such Encyclopedia Britannica and Microsoft Encarta, but they were static encyclopedias where the users were limited to a consuming role, not being able to create lemma's or correct errors. User Generated Content is a broad term for media content created by and for end-users (Kaplan, Haenlein, 2010:61). User Generated Content has to fulfil three criteria: it needs to be content published on a publicly accessible website, or on a social networking site available to a select group of people, it needs to show an amount of creative effort and it needs to be created outside professional routines and practices (Kaplan, Haenlein 2010:61). With the spread of the web 2.0 concept and several technological developments, combined with generations of children growing up with the internet, this definition seems solid: in order to be a social medium the social
way of thinking prevalent on the internet these days must be used.

But what about Social Media that do not use Web 2.0 characteristics? Are sites that share information and allow user input without using web 2.0 terms and techniques unsocial? Much like the well known “sender – message – receiver” model of Shannon and Weaver (1949) the definition of Kaplan and Haenlein (2010) is focused on the technical realisation, rather than on the use. Using Kaplan and Haenlein's strict definition, it is not hard to imagine an example of Social Media that falls outside their scope. Imagine for example a forum where users share their own written tales set in the fictional universe of an original, professionally written work of fiction, so called fan fiction (Wikipedia f, 2015). It receives its content from its users, and due to the nature of online fora allows for easy sharing of the user created material and reactions on said material by other users. A forum of this sort is fully realisable without any Web 2.0 characteristics as for example Flash or AJAX, a requisite according to Kaplan and Haenlein (2010) . Or of course Reddit, originally written in LISP, a programming language that has been around since 1958, and now in Python (Reddit c, 2005, Wikipedia b, 2015) is another example of a social medium that falls outside the strict, but widespread definition of Kaplan and Haenlein (2010).

The ideological concept of Web 2.0, the sharing of content and the influence of the end user, do not necessarily mean that all the buzzwords and buzz-techniques of web 2.0 are a requirement for a Social Medium. Lietsala and Sirkunnen (2008: pg.13) provide a broader definition of Social Media: “a term that is used to describe web services that receive most of their content from users or that aggregate the content from other sides as feeds”. Social Media is more of an umbrella term given this definition (Lietsala and Sirkunnen 2008, pg. 17) and fully applicable to Reddit.

2.2 Describing Reddit

In this paragraph we will describe Reddit before we cover the concept of repairs and the theoretical background, beginning with a description of the front page of Reddit (2.2.1), followed by an explanation of the way the posts are ordered (2.2.2). Then, we describe the concepts of down- and upvoting (2.2.3) followed by a description of some of the rules Reddit expects users to adhere to (2.2.4). Finally, we will cover the way a redditor can edit his or her post (2.2.5).

2.2.1 Front page and subreddits.

Reddit is not the most hip looking website on the Internet and at first glance its homepage, or the front page as it is commonly known on Reddit is a nigh incomprehensible screen filled with links and text spanning almost the entire width of the screen. On this front the most eye-catching posts on
the whole of Reddit are singled out. These can be direct links to for example image hosts, online newspapers, Youtube or posts within Reddit itself.

Rather than being an uncategorised hotchpotch of submitted content, Reddit enables users to start their own sub-forums (subreddits) within the main Reddit site. Subreddits can be about any imaginable subject. These subreddits have a specific URL form: www.reddit.com/r/nameofthesubreddit. The short form /r/nameofthesubreddit is used across Reddit to indicate a subreddit, we will use that form here as well too. In this subreddit a registered user can create a post, and other registered users can respond to that post. Users can also reply to responses. For non registered viewers, Reddit is read-only; although it can be read, viewers can not respond to or submit content. Subreddits are moderated and posters are expected to stay on topic within a subreddit. Many subreddits have their own set of rules of use, in addition to Reddits general rules for example a ban on post that are pornographic in nature.

In effect, users of Reddit (redditors) create a tree structure of replies to a mother post. These tree structures have several interesting communicative features that could be researched in the future. Reddit has also developed its own jargon. Some examples of that are seen in the screen shot below: IamA (I am a) is an abbreviation that at given the context, or even phonetically is
understandable for new viewers, but AMA (Ask Me Anything) isn't. There are other examples of specific jargon for Reddit, as well as the use of general Internet-speak jargon, however we will not cover those here unless needed.

In the screen shot below we can also see the tree structure of the replies to the original post. The replies to replies are indented for readability.

![Screenshot of Reddit AMA](image)

Picture 2. The replies to a post are indented. Note that replies can be indented even deeper than displayed here.

### 2.2.2 Determining the order of the posts

When a post is made in a subreddit, in general it is added to the already existing posts in that subreddit. These posts can be sorted in a variety of ways, as can the replies to a post. By default the sorting is by what is “hot”, the most interesting posts are placed at the top of the page. The same goes for the replies to a post or to other replies. There are other orders of sorting, for example in chronological order, however as these orders are not used in this research we will not cover them here.

Determining what is hot is done by an algorithm that takes the time of the post into account, as well as the number of upvotes and downvotes by the redditors (Salihefendic, 2015). The better the score, the hotter the post or reply is and the higher it gets displayed. This sorting order can be applied (and
2.2.3 Upvoting and downvoting

In determining the order of posts the term upvoting and downvoting were mentioned. Reddit’s upvote/downvote system works as follows; redditors (whilst logged in) can upvote and downvote posts by clicking on either an arrow pointing up or down on the left side of a post. An upvote gives a point to a post, a downvote takes points away. Posts with too many downvotes are hidden from view, and require a click on a “load more comments” button to show. These terms are not unique for Reddit, and are implemented on other sites sometimes with different terms. Sites can for example implement a “kudo” system where users give kudos (points) to posts they deem good, and take them away for posts they deem bad.

2.2.4 Rules and Reddiquette

As with many sites that rely on content from users there are certain rules that users have to follow. Reddit, citing free speech, allows an extraordinary variety of subreddits including some that are pornographic or blatantly racist in nature but has and enforces a set or rules (Wikipedia a, 2015). Child pornography, spam and the posting of personal information are prohibited (Reddit a, 2015). As said above, many subreddits have extra rules in place to have more control over the content of their subreddit: for example not allowing content that is deemed “not safe for work” (NSFW).

Additionally to the rules, Reddit also relies on “reddiquette” (a portmanteau of Reddit and (n)etiquette) to keep Reddit civil. Rather than rules the reddiquette is formed as polite suggestions and users are asked to abide by them (Reddit b, 2015). Users are asked not to troll, insult, be rude, ask for votes, make pointless comments and more, but are also asked to remember that there is a human being on the other side of the line, use proper grammar and spelling, and read the rules of a subreddit (Reddit b, 2015).

The reddiquette also gives the following suggestion: “State the reason for any editing of posts”. In face-to-face conversation there hardly is a need to indicate a reason for a repair this explicitly: in general it can be logically deduced from the conversation. However, in online conversation the reason for an edit or a repair can be more hidden. Reddit not only asks users to give a reason for their edit, but also marks the edited post with an asterisk and changes a value in the meta data. This will prove to be invaluable in the analysing of repairs later on.
2.2.5 Editing a post on Reddit

Reddit allows redditors to edit their own posts. Logged in, a redditor is able to edit his or her own posts. By clicking on a simple link underneath a post, a redditor can edit his post via a text box that pops up in place of the original post.

Picture 3. A logged in user (basvandermeer) can edit his own post by clicking the edit link.

The subject of the above screen shot is about how to indicate the strips (playing field) for a fencing tournament. In one of the scarce posts made on Reddit by the identifiable account used for this thesis I suggest the use of chalk. Both in the posts above and below mine, the edit button isn’t visible. Pressing the edit button (here underlined in red for clarity) changes the image to the following:
Approximate. However, reels move. They would need tape. Also, no grounded strips. This is a very large tournament in a gym at a university, not a salle, or anywhere where grounded strips could be conceivably shipped in.

(...and anyway, all of those cam-lock strips are probably headed out to the Foul World Cup in SF for the same weekend.)

And it would not be possible to use chalk to lay out the strips?

It’s varnished floor. They can’t tape anything to it (paper for strips, taping down reels, etc.). Could they do chalk lines? Sure. If they want to have one on-hand for EVERY single point, with two people assigned to each strip just to re-snap the line every time a phrase ends.

Would I do it if that was the absolute LAST option? Personally, yes, but professionally, I’m still concerned that the chalk would make it slick, not to mention a royal pain in the ass to ensure that the lines are “legal”.

This allows me to edit the post. Note the links to both the reddiquette and help on formatting text on the right. It is in this text box that following reddiquette the reason for an edit can be given.

I recently went to Auschwitz and apparently if you were to run away, at least 10 people were killed for every person who escaped. Your family were the first to go if they were around.

So I think that’s what he meant ... I could be wrong but yeah.

EDIT: Spelling.

POW camps designed for soldiers were not the same as concentration camps in the slightest, especially on the western front. The western front followed the old rules of war and they treated soldiers, generally, with the utmost respect.

OK, so my statement was a little bit too general and obviously does not apply to all during the second world war and my statement was generally regarding allied troops (read UK/US/CAN/AUS/FR) that were captured by Germany. Obviously those troops from the eastern front did not have such treatment and some games for those in the pacific theatre.

Between certain countries they were and certain countries they weren’t. As someone has already pointed out Nazi’s and Soviets weren’t very fond of each other and Japanese felt that soldiers who surrendered were a disgrace. The Japanese didn’t abide by the Geneva Convention.

I had an ‘uncle’ who was a POW in a Japanese camp. He didn’t say much and we’re talking back a couple of decades ago when you neither asked or said anything. Sad I know but I was just a kid. Only thing I heard was him hopping rifle butts in the back.

On another note, was the Cowra breakout. I went there a couple of years ago and learned after the visit that the German and Italian POW’s didn’t join in the breakout. They were even allowed trips to town and some of them even stayed after the war. Cant blame them considering it is a nice bit of country down that way. You would be mad not to.
In the above screen shot there are two examples of repair adhering to the reddiquette. In both cases the redditor gave a reason for his or her edit. Once for spelling and once to clarify a statement.

2.3 What is a repair?

In this paragraph we will begin by giving a definition of what a repair is (2.3.1). Followed by a short sideline to the way repairs are created in the brain (2.3.2)

2.3.1 Repairs

As the focus of this thesis lies on the use of repairs, before delving into anything else, it makes sense to start by giving a definition of what a repair is. Communication is inherently fuzzy: it always possible for a message to be wrongly interpreted by the receiver or wrongly formulated by the speaker. In both cases some action is required in order to be successfully understood. Levinson (1983) defines a repair as an alteration that is suggested or made by a speaker, the addressee, or audience in order to correct or clarify a previous conversational contribution. Repairs may occur at any of several points following the contribution in question, occurring in accordance with a conventional order of preference (SIL 2014, Levinson 1983). There are several forms of repair: self-initiated self repairs, in which a speaker repairs his own utterance before the addressee can intervene, other-initiated self repair, here the addressee asks for clarification and the speaker then repairs his utterance and finally other-repair where the addressee corrects the speaker (Levinson 1983). We will covers these forms in more detail below. In speaking, the preferred form of correction, or repair is the self-repair (Schegloff ea 1977, Levinson 1983).

We will use Levelts classification of repairs as a starting point, as his system is more focussed on the repair itself, rather than the context of the repairs. This provides us with a neutral starting point which we will expand on. We will go into detail of both Levelts system (in 2.3.4) and our additions (in 2.6) below.

Repairing in text based communication is rarer than in face-to-face communication. Analysing and researching repairs is primarily focused on spoken interaction; repairs can be well heard, researched and are relatively common (Jacobs & Garcia, 2013). Preliminary observation shows that it is relatively easy to find examples of repairs in Reddit.
2.3.2 Creating and monitoring of repairs

There is an ongoing debate regarding the way the user monitors his utterances and finds the need to create repairs. Perhaps a bit of a side-line, but as these developments are quite interesting and may be of use in future research we will nevertheless mention these. The standard theory regarding the way a speakers controls whether or not his utterances need repair is the perceptual theory. The perceptual theory of monitoring by Levelt states that a speaker monitors or checks his inner and outer speech and via that path finds and corrects his errors (Levelt 1983 via van Hest 1996). This theory, however, is subject to debate, and although the debate is focused on monitoring speech (in whatever way or form), the way utterances are checked for errors may be of influence on written discourse such as Reddit. As such, we will briefly cover one of the papers critical of the perceptual theory.

Nozari, Dell and Schwartz (2011) state that there is more to repairs than originally thought by Levelt. Levelt has stated that utterances are continuously monitored for errors by the so called perceptual-loop account. On the assumption that speakers cannot monitor their own speech production for errors, the detection of errors is based on the monitoring of the output by the speech comprehension system. (Levelt 1989 via van Hest 1996) Essentially: “speakers detect errors in their speech by listening to themselves” (Nozari, Dell, Schwartz, 2011). Nozari, Dell and Schwartz (2011) state however that this monitoring of the external channel does not provide sufficient explanation for all findings of error detection. They base this claim on the measuring of the latency time between the detection of the error and the initiation of correction. In short: the time between error detection and cut-off was too short when compared to the time it takes to recognise a word and halt the ongoing action (Nozari, Dell, Schwartz, 2011). This discrepancy was solved by the monitoring of not only the external speech, but also the internal speech by the comprehension system. Using data gathered from subjects with aphasia, a strong correlation between error detection ability and the model's characterisation of their production skills was found, supporting an internal conflict-based monitor (Nozari, Dell, Schwartz, 2011). This conflict based monitor is based on the premise that: “an error is signaled (sic.) by the presence of high conflict between various options at the time of selection among activated words, and later, among activated phonemes.”(Nozari, Dell, Schwartz, 2011:25). Even though they provide a new model for error detection Nozarri, Dell and Schwartz (2011, p. 26) do not “deny the role of perceptual processes in speech monitoring altogether”.

An interesting discovery that may or may not be applicable to error detection in Reddit, and the subsequent repairs. As redditors create content on Reddit, they have several points on which
they can monitor their utterance (so to say) for errors: when thinking about what to write, when actually creating their post, and during the reading of their post, either before or after they post it to Reddit. It may very well be that when checking for errors in written discourse a perceptual monitoring system is used only after the full post is written and posted. However, this is food for thought for further and more specialized research beyond the scope of this thesis. At this point it suffices to say that errors are being made, detected by the redditor or others and repaired by the redditor via the editing of the original post.

2.3.3 Types and places of repair

Repairs and the errors being the cause of a repair can be classified in many ways. As one of the “founding fathers” and one of the first to classify and categorise errors and subsequent repairs Levelt (1983) developed a system that was subsequently implemented in one form or another by many other researchers (see for example: Postma & Kolk, 1990, Plug 2011, van Hest 1996 and Smith 2009). Levelt (1983) first makes a primary distinction between covert and overt repairs and categorises the latter further. In the paragraphs below we will cover the categorisation by Levelt and (adaptations by) others and see if and how this categorisation can be applied to Reddit.

2.3.4 Covert repairs

Before looking at examples of repairs that are visible, via transcript or otherwise, so called hidden or covert repairs must be pointed out. Levelt classified repairs as either covert or overt (Levelt 1983 via van Hest 1996). Overt repairs are made after the erroneous utterance has been made. Whereas repairs that are hidden, or covert are not explicitly spoken but made by the speaker before speaking the utterance (Plug 2011). Hidden repairs can take the shape of hesitations, repetitions and restarts (Plug 2011). Examples could be sentences such as: “eeehm, could you pass, pass the sa...salt”. This is an artificial example, but it is not hard to find examples in everyday interaction. Schegloff (2013) however states that restarts, or recycling of a part of a turn constructional unit (TCU), is an operation in its own right and can, for example, be used to frame a repair, or as a way to (re) commence a turn if there was some sort of overlap. Schegloff (2013) gives, amongst others, the following example:

06 Bee:    →rcl  I don’know. The school– school uh, (1.0) bookstore
07                 doesn’ carry anything anymo (h) uh,
Here the recycling-action of the repair is indicated by “--\(\rightarrow\)rcl”. Even though the debate is ongoing, it is safe to say that repairs that aren’t (overtly) explicit need to be categorised as covert repairs. There are many points in time during the composing of his or her message in which a redditor can covertly repair his message, for example when the message is being written but not yet posted to Reddit. With the research method used in this thesis we can only observe the repairs made by editing a post after it has been placed on Reddit. We will cover the notion of covert and overt, or private and public repairs in more detail below.

2.3.5 Overt repairs

In contrast to covert repairs, where the repair itself isn’t visible to anyone but the speaker, overt repairs are repairs made “in full view” of the listener(s). Levelt found that repairs tend to take a distinct shape (Levelt 1983 via van Hest 1996). Overt repairs are made after the error in the utterance has been pronounced (Levelt 1983). These repairs contain three distinct parts: the original utterance containing the part that needs to be repaired (reparandum), the cut-off point after which an editing phase begins. This editing phase also contains the editing term, such as “a”, “uh”, “well” et cetera. The repair ends with the reparatum, a proper and corrected utterance. (Levelt 1983, van Hest 1996). Overt repairs can be and have been categorised further, Levelt’s (1983) categorisation has been used as a starting point by other authors (Postma & Kolk, 1990)(Plug 2011) or adapted, for example when researching second language acquisition (van Hest 1996), (Smith 2009). Given the importance of his system and the distinct focus on the repairs itself, rather than incorporating repairs in the context of the conversation, Levelts (1983) categorisation and examples are used as a starting point here as well, and described below. As we do not yet know the way repairs are being used in a conversation on Reddit, it is probably best to begin by focussing on the repair first, and only later on its role in the conversation.

Speakers may change their mind regarding a message they are currently speaking. Perhaps because the new message is easier to say, or could be better comprehended by the listener. The speaker consequently replaces his original message with a different one. Levelt (1983) calls these repairs D-repairs and states that they are infrequently found.
Besides D-repairs Levelt also distinguishes several varieties of *appropriateness-repairs* (*A-repairs*). In these repairs the speaker is concerned about whether or not his message is understood given the context of his utterance. There could be potential ambiguity given the context, Levelt gives the following example:

"We beginnen in het midden met . . . , in het midden van het papier"

*We start in the middle with . . . , in the middle of the paper*

"met een blauw rondje"

*with a blue disc.*

(Levelt 1983, pp 52)

Here "*in the middle*” could have two different meanings, so the speaker reduces the ambiguity of his statement by adding the "in the middle of the paper” to his utterance. Repairs like these are labelled *ambiguity reduction repairs* by Levelt. Other examples of ambiguity reduction are utterances that repair unclear demonstratives (a demonstrative is a determiner that is used deictically to indicate a referent’s spatial, temporal, or discourse location (LingualLinks 2014)).

Besides reduction of ambiguity, appropriate levels of describing a concept can also be cause for repair. Levelt (1983, pp 52) gives the example of a *blue spot* becoming a *blue disc* and states that repairs of this type tend to go from less detailed to more detailed and calls these repairs *appropriate level repairs*.

It may also happen that a speaker needs to repair his utterance in order to make a coherent as possible statement given his previous, or perhaps even first part of the utterance. Levelt (1983, pp 53) uses the example

"ga je een naar boven, is... uh kom je bij geel"

*"go you one up, is uh... come you to yellow”*

Here, although the sentence could be finished by making a noun-phrase (NP) with the verb “is”, it makes more sense, in Dutch grammar at least, to finish the sentence with a NP containing the verb “komen” (to come, or to arrive at). As the originally intended utterance would be coherent, Levelt doesn’t classify this as an error but rather as *ambiguous coherent repair*, and mentions that distinguishing between a speaker repairing the ambiguity level or establishing coherence isn’t always unambiguously possible.

Besides repairing ambiguity, it may happen that a speaker discover that what he is saying contains some sort of error. More straightforward than ambiguity repairs, *error repairs* simply
repair errors made by speakers. Levelt classifies the following types of error.

First there are *lexical errors*. Here the speaker constructs a sentence, discovers that an incorrect word is used and repairs the utterance. Lexical errors can be present in almost any part of the utterance. For example:

(7) Linksaf naar,…. rechtsaf herstel naar blauw

*Left to, …. right correction to blue*

Errors of this kind were amongst the most common in Levelt's study.

Besides lexical errors Levelt names *syntactic errors*, in which a speaker syntactic construction isn’t capable of construction a proper utterance, so the speaker corrects himself. Levelt (1983, pp 54) gives the following example:

En zwart . . . van zwart naar rechts naar rood

*And black . . . from black to right to red*

Levelt found that these errors were not very common in his study. Finally Levelt covers *phonic errors*, these aren’t relevant for this thesis but are errors in speech where the speaker uses the wrong phonemes to create a word: “unut, unit”.

Finally, Levelt classifies all repairs that don’t conform to any of the aforementioned categories as “rest”. Further on in this thesis we will come back to Levelt's classification of repairs and apply his model to the repairs found in the Reddit source material.

Levelt's categorisation for repairs is based on the first language, the mother tongue of a speaker. Hence the examples in Dutch, with literal English translations. Repairs are therefore prevalent in the original language of a speaker, but making errors and repairing them is not restricted to the speakers first language. Repairs are also prevalent when learning and speaking a second language. For example Green and Hecht (1993) found that pupils who were allowed to make errors and correct them themselves performed better in learning English as a second language. See also Kormos 1999 for more research on self-repair in a second language. As it is perfectly possible that not all users of Reddit have English as their primary language, it may be necessary to add a second *language-error repair category* to Levelt's model, provided the nationality of the users can be determined.

On an interesting side note: repairs needn’t solely be made by humans. Moore, Churchill and Kantamneni (2011) researched repairs in interaction with an internet search engine. Their findings included the notion that people using the search engine can and do repair their original queries if the responses offered by the search engine do not provide an adequate answer but also that the suggestions given by the search engine can have the same function as a repair. Seeing the interaction with a search engine as a dialogue, complete with repairs coming from both the search
engine and user may improve the effectiveness of search engines (Moore, Churchill and Kantamneni, 2011).

### 2.3.6 Self-initiated and other-initiated repairs

Where Levelt's focus is solely on the repair itself, repairs can also be analysed according to their place in the conversation and by the person initiating the repair. We have already mentioned the self-initiated self-repair, the other-initiated self-repair, the self-initiated other-repair and the other-initiated other-repair and will cover these in more detail below.

Before going into these repairs, we must at least briefly cover the concept of turn-taking and adjacency pairs. Turn taking in a conversation means nothing more than that participants switch their talking and being silent in an orderly fashion. A local management system, with a set of rules allows participants to take control of the “floor”. Initially a participant will have a *turn-constructional unit* assigned to him, allowing the initiation of a turn, at the end a turn has a *transition relevance place*. The most obvious example of turn taking and transition is a question, the question itself is an turn, taking the floor. After the question is asked, the turn is finished and other participants may take the floor and provide an answer in their turn (Levinson 1983).

Besides turns, there are adjacency pairs. Paired utterances such as the question-answer, offer-acceptance etcetera. These adjacency pairs can have sequences inserted in between. Adjacency pairs have preference organisation, that is to say that although there are a myriad of possible replies to the first part of an adjacency pair, there is a preferred one. Preferred here does not point to the preference of the speaker or listener but rather to the most unmarked, most simple and natural feeling reply. Preference can operate within and over turns integrating with repairs. A listener may feel disinclined to answer with a dis-preferred reply and may use repair as a way to allow the speaker to reformulate his message (Levinson 1983). Additionally, repairs are a “device for the correction of misunderstandings, mishearings or indeed non-hearings” (Levinson 1983: 340).

Levinson provides a description the properties of repairs that differs than that from Levelt. Rather than focussing on the precise content of the repair, Levinson aims to explain the concept of repairs within a conversation. As a device, repairs provide four slots in a (at least three-turn sequence) for the actual repair, or the prompting for a repair. The first T1, includes the repairable item and is the first opportunity for self-initiated self-repair. In the transition space between T1 and T2, there is a second opportunity for self-initiated self-repair. In T2, there is the third opportunity for repair, either for an other-initiated other repair or an other-initiated self-repair. In T3, the last opportunity, if there was an other-initiation in T2 there is an opportunity for an other-initiated self-
repair. The preference in uttering a repair goes from a self-initiated self-repair in T1 to an other-initiated other-repair in T3.

What does the preference for repair have to do with Reddit? When keeping in mind that redditors have ample opportunities for repairs before their post is displayed online, it is curious to say the least that examples of edited posts containing one or more repairs are so easily found. A written message, carefully constructed would not need a repair, would it? Could one say that the conversation, and thus the mechanism for repairs, starts when the post is made public, comparable with an utterance made up in the mind of the speaker and only part of the conversation when spoken? Without finding evidence to the contrary, we will assume this to be mostly the case. Small spelling errors, reformulations by re-writing a sentence etcetera could all be corrected before posting, but repairs based on the interaction with others cannot be accounted for in the making of the message. Redditors aren’t mind readers and posts can be posted simultaneously, resulting in an unintended order. Reddit does have moderators tasked with ensuring the rules of a subreddit are followed, but we were only able to find one example where a moderator actively edited the posts of others\(^1\). Such an action can be classified as an other-initiated other-repair but even then we are stretching the definition as the post is edited in such a way that neither the original post, nor a post after a repair remained.

Therefore, for the time being we assume that all edits to a post are made by the original poster, and are thus either self-initiated self-repairs or other-initiated self-repairs.

\(^1\) https://www.reddit.com/r/AskHistorians/comments/1svj1q/how_did_people_esp_european_townsmen_get_fresh/
2.3.7 Repair after next turn

Assuming all edits to a post are made by the original poster, whether or not they are initiated by others, we need to cover at what point in the conversation the edits containing repairs (in spoken interaction) are made. It seems obvious that an edit can be made after a listener or other redditor replied to the initial post. However, this is a more complex phenomenon that needs in depth explaining. Schegloff (1992) provides us with insight. In *Repair after next turn: the last structurally provided defense of intersubjectivity in conversation* Schegloff covers the use of the third position repair as a defence for the idea of intersubjectivity, that is the notion that the discussed subject is understood similarly by the participants in the conversation. Colloquially said: not only are they talking about the same thing, they are thinking similarly about said thing. Looking at an example from Schegloff (1992) below we see that even after a speaker has spoken and has been replied to he can still repair his utterance, and does so in a very consequent manner. This in contrast to Levelt's examples of repair mentioned above, where repairing often takes places in the same sentence.

Another item mentioned by Schegloff (1992) is that third position repairs can have turns between the trouble source and the repair. Whilst still being the third position sequentially, there is a possibility that turns are interjected. Even a repair in the fourth position, referring to a trouble-source in the first is possible. More so: Schegloff mentions “ongoing multiple repair spaces” and provides us with a schedule displaying a myriad of possible third position repair spaces. In the examination of the example below Schegloff (1992:1322) explicitly mentions the displacing of the forth position repair from the serially fourth turn. Here both father and son are named Phil.

Excerpt 21 (CDHQ, 15; Openings, 299)

| Phil: | Hello? |
| Lehroff: | Phil! |
| Phil: | Yeh. |
| Lehroff: | Josh Lehroff. |
| Phil: | Yeh. |
| Lehroff: | Ah:: what've you gotten so far. Any requests to dispatch any trucks in any areas, |
| Phil: | Oh you want my daddy. |
| Lehroff: | Yeah, Phil, |
| Phil: | Well he's outta town at a convention. |

What does this mean for Reddit? In the paragraph above, we stated that the preference for a repair at the third position is one of other-initiated other-repair, yet apparently this notion is
contradicted by this research. Schegloff’s (1992) research was done on spoken interaction, so a one-on-one comparison to written CMC, such as Reddit will run aground. Nevertheless, Reddit provides ample opportunities for repair prior to the posting of the message, and the opportunity to edit the message after it has been posted (possibly to repair part of the message). The question at what position a redditor makes the choice to edit and/or repair his message is an interesting one. Redditors might consistently edit their post to repair their post, in response to a new, different, turn taken by some other redditor, rather than repairing in the same turn. To answer that question, an analysis of the conversation within one or several threads has to be made. However, in this thesis we aim to provide a broader overview of the use of repairs on Reddit, rather than an in-depth case-study. We can only acknowledge the fact that repairs can refer to a trouble-source quite some turns away from the the repair and must hope to research this matter at some point in the future. The problem of turns being inserted between what Schegloff calls the “trouble-source” and the repair is however something we can and must cover in this thesis. As we look at repairs were the redditor explicitly stated his reason for a repair, there is a possibility that the reason for his repair is a post made by another redditor somewhere in the thread. Additionally, we have to acknowledge that position and sequence on Reddit aren’t always at the same time and place in the conversation; there is ample room for displacement on Reddit. The concept of displacement is covered in more detail in 2.4.6.

2.3.8 Private versus public editing in CMC/Reddit

The example of a moderator editing a post of another redditor is both an example of an other-initiated other-repair and of a public edit. There are forms of CMC that allow a user to change or even completely remove his or her message, whereas other forms of CMC do not offer such an option. Take for example e-mail; once the mail is composed and send it will receive the recipient as is. Only recently did Google build in a delay in its Gmail application that allows users to un-send a message. In IRC-chat, once a user presses send the constructed message is visible for all other users and there is no possibility to edit that displayed message. All editing and repairing before a message is done in private, invisible for other users.

On the other hand, online forums, Facebook and also Reddit allow a user to change a message after it has been posted. Facebook shows the original message, as do some forums. Others show what percentage of the message has changed or in the case of Reddit display an asterisk to indicate that the message has been altered. Repairs done on a message visible to others are said to be publicly edited.
2.3.9 Other uses of repairs

Speaker meaning, the communicative intention of a speaker about the meaning of an utterance in context (Sanders 2013) may also be cause for the use of a repair. However, whereas repairs are usually seen (for example above) as a correction of an error of some kind, provides a different reason for repairs in conversation. In *The duality of speaker meaning. What makes self-repair, insincerity, and sarcasm possible* Sanders (2013) distinguishes between personal intentions and generic intentions when it comes to assigning meaning to an utterance. A speaker may have a personal intention with his utterance that differs from the meaning generally attributed to that utterance. Sanders give an example regarding the behaviour of a dog:

(5) Casey has been. . . well, pleasant.

and explains that although there are many ways to describe the behaviour of a dog, pleasant generally isn’t one of them, indicating that although in general it would seem that there was nothing wrong with the behaviour of the dog, in this specific instance there most likely was in the eye of the speaker. When looking at self-initiated self repair Sanders (2013) states that such a repair can be used to align the generic intention with the personal intention. Here he uses an example of a break up between (romantic) partners via telephone. When reading the excerpt from the transcript it seems that the caller breaking up with his partner is sincere, however Sanders adds the following disclaimer:

“[…] *we do not know whether the speaker is sincere and wants to be credited with that, or insincere and wants to conceal it. Either way, through the succession of utterances he produces, each arguably repairing a defect in a prior utterance, the speaker seems to make a discursive effort to be credited with actually intending_{pers} what a generic speaker would intend_{gen}, the speaker meaning of his utterances to be.*” It seems that being insincere, asserting sincerity or being outright sarcastic can be done via repairs. It could be that repairs are used this way in Reddit. A user might for example say something along the following lines: “*what a great idea, oh wait!*” to indicate sarcasm and to underline his intended sarcasm even further do so not by writing it in his original post, but by repairing his post to add the sarcastic meaning. But it is the idea that repairs can be used for a purpose other than clarification or to correct an error that is most important here. Instead of strictly defining that a repair is only for correction or clarification, being able to use repairs to convey a variety of other intentions allows speakers a far more complex use of repairs.
2.3.10 Distinguishing between an edit and repair on Reddit.

Reddit and other forms of CMC allow the editing of a post after it has been made. A user can use the edit function of his chosen form of CMC to repair an error found in his post, but a post might be edited for reasons other than repair. A post might have broken one of the “house-rules” of a subreddit and be made blank (but not deleted) by a moderator. That is certainly an edit, but not a repair.

So, while it is possible to simply equal an edit to a repair and use the terms interchangeably, it is better to at least acknowledge that this doesn’t always has to be the case. That is, the editing of a post on Reddit doesn’t have to confirm to a repair as seen in the face-to-face examples mentioned above. We did however observe edits that can be seen as a repair, but don’t adhere to the categories of repair as classified by Levelt (1983) and Schegloff (2013). And as seen in the above paragraph, a repair can be or mean more than a simple correction of an error. In this thesis we will make the following distinction between repairs and edits. We will strive to use the term repair for that part of a post that is an actual repair as laid out by Levelt aiming to correct an error or provide clarification, or if said part confirms to the idea of Sanders (2013) and is a repair with a more complex underlying intention. So we aim to use the term “repair” and so indicate the speech act and use the term “edit” to indiscriminately indicate the making of any alteration to a post on Reddit. Therefore an edit may contain a repair, although this reversibly doesn’t need to be the case. For example, the sentence “we shall walk, no run home” contains a repair, but could be posted to Reddit without making an edit to the post containing that sentence.

2.4 From the collaborative model of communication to grounding in CMC

Having covered both Reddit and repairs we move towards the communication model used in this thesis. Beginning with the maxims of Grice and his introduction of the collaborative principle (2.4.1) followed by the collaborative model of Clark and Wilkes-Gibbs. We do this to provide a framework in which to place this thesis and to explain and introduce jargon used here. After all, it is one thing to state that repairs correct errors, but without setting the framework and providing a context as to why it is important for the success of the communication this thesis would be pointless.

Even when critically analysing and dismissing Shannon and Weavers model from the
introduction, there is an undeniable truth in Weavers statement that if communication already fails on a technical level, the understanding of the message and if necessary acting on the conveyed message will quite possibly fail too (Shannon, Weaver 1949 pg. 6). But what if the communication succeeds on a technical level, how do we know whether or not we have communicated successfully? Are there rules to successful communication? The cooperative model of communication allows us to formulate satisfying answers to these questions.

2.4.1 Grice’s maxims

In 1975 Paul Grice lays out what would come to be known as Grice’s maxims and the cooperative principle. These maxims serve as the foundation for the collaborative model, a model of communication focused on the interaction and collaboration needed between partners to communicate successfully.

Grice points out that conversation isn’t a set of “disconnected remarks”, but rather a cooperative effort in which every participant sees a common purpose or “at least a mutually accepted direction” (Grice 1975: 45). Grice then states a general principle which participants are expected to observe and calls this the cooperative principle: “Make your conversational contribution such as 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, 1975:45). To adhere to this cooperative principle, participants must follow four maxims pertaining to quantity, quality, relation and manner as set out in the table below.

Table 0: The maxims of Grice

| Quantity         | 1. Make your contribution as informative as is required (for the current purpose of the exchange).  
|                  | 2. Do not make your contribution more informative than is required.  
| Quality          | A supermaxim: try to make your contribution one that is true.  
|                  | 1. Do not say what you believe to be false.  
|                  | 2. Do not say that for which you lack adequate evidence.  
| Relation         | 1. Be relevant  
| Manner           | A supermaxim: be perspicuous.  
|                  | 1. Avoid obscurity of expression.  
|                  | 2. Avoid ambiguity.  
|                  | 3. Be brief (avoid unnecessary prolixity).  

26
Grices Maxims and the cooperative principle point out that communication is more than a simple exchange of messages. It is this notion that serves as a foundation for the collaborative model. In this model repairs can be made for more than just the reason that a message didn’t reach its intended recipient. One could make a repair to a previously made utterance after it has been proven untrue for example or to clarify an (unintentionally) ambiguous sentence, as also seen in Levelt (1983).

2.4.2 From the cooperative principle to the collaborative model

From Grices initial observations we move on to the formalisation his maxims (amongst other observations) in the form of the collaborative model by Clark and Wilkes-Gibbs. The collaborative model of communication is a reaction against what Clark and Wilkes-Gibbs call the literary model (Clark, Wilkes-Gibbs, 1986:3). In this model “speakers refer as if they were writing to distant readers” (Clark, Wilkes-Gibbs, 1986:3) The literary model is a simplified model with the eye-catching assumption that the speaker alone controls the conversation. (Clark, Wilkes-Gibbs, 1986:3)

When analysing conversation however, this model simply isn’t adequate, it is oversimplified and not all appropriate for conversation. Written messages can be read, re-read and edited to the writers delight, whereas speakers only have a limited time to construct a message, written messages (in general at least) can be re-read, not only by the writer but also by the addressee. Speech is a process of several actions combined: listening, uttering, understanding are all done at, almost, the same time in order to understand the intended message. Finally, where in written communication the actions (or lack thereof) of the addressee do not matter, the listener matters in the conversation; he or she might express a lack of understanding or interrupt or do a plethora of other actions that can all influence the way the speaker needs to construct his message (Clark, Wilkes-Gibbs, 1986:3). Although Reddit is a written medium, we've observed that interaction on Reddit can't properly be described as the back-and-forth sending of written letters. Interaction can closely resemble online chat, where messages follow each other in rapid succession, mimicking spoken interaction.

What then defines the collaborative model? The defining feature of the collaborative model is the idea of mutual acceptance. Both speaker A and listener B must mutually believe that B understood A's references in order for the conversation to continue (Clark, Wilkes-Gibbs, 1986:9). This process, an acceptance process is based on a presentation and an acceptance phase. Speaker A presents a phrase to listener B, assuming not only that B is able to listen and understand A, but also
that B is capable of picking out the referent based on A's description. Having presented his
description, A now waits for B to either accept or not accept A's “presentation” (Clark, Wilkes-
Gibbs, 1986:9). Accepting can take an overt form, where B asserts acceptance by using continuers
such as “yes”, “right”, “I see”, etc. or even a simple head nod. B can also presuppose acceptance;
either by moving on to the next contribution or by allowing A to continue. Both are seen as
acceptances of the last made presentation (Clark, Wilkes-Gibbs 1986:9). Using this collaborative
model, in time the same amount of information can be transferred with less expense on behalf of the
speaker (Clark, Wilkes-Gibbs, 1986). The description above applies to exchanges between A and B
where A's utterance was understandable and acceptable by B.

What if the initial utterance isn’t acceptable? Clark and Wilkes-Gibbs (1986) name three
ways to refashion the initial utterance into an acceptable one. The most important one for this
thesis: the use of both covert and overt and self and other-initiated repairs (Clark, Wilkes-Gibbs,
1986:21). In their experiment Clark and Wilkes-Gibbs (1986) found that refashions were
(relatively) prominent at the beginning of the experiment, but barely present later on. Classical
theories of least effort state that “speakers try to utter the shortest noun phrases that will enable their
addressees to pick out the referent in context” (Clark, Wilkes-Gibbs, 1986:26). Clark and Wilkes-
Gibbs proposed that such a minimising of effort also exist in the collaborative model. Adding to
that, there is a trade-off between the effort a speaker puts in his initial noun-phrase, and the
refashioning of it. The more effort put in the first, the less need there (likely) is for refashioning
(Clark, Wilkes-Gibbs, 1986:26).

This gives the use of repairs in the conversation a place in the collaborative model. By using
repairs it is not only possible for speakers to correct errors as they make them (for example as a
self-initiated self-repair) but also by the listener to intervene by repair in order to quickly minimise
the misunderstanding. In both cases, using repairs in the collaborative model enables the
participants to minimise the number of exchanges needed before mutual acceptance (Clark, Wilkes-
Gibbs, 1986:27). Clark and Wilkes-Gibbs proved that it's the collaborative model of communication
that requires that interaction between participants is needed in order achieve mutual understanding.
This interaction may go perfectly and require no refashioning, but when there is need for refashion
a repair may be in order. The collaborative model, the principle of least collaborative effort, the use
of refashioning and repairing to minimise misunderstanding all point to the participants will to
create what is essentially a common ground between them. Redditors may have difficulty
understanding each other, in the broadest sense of the word. To create mutual understanding
redditors may simply place a follow-up post explaining themselves or use the repair function to
explain within the misunderstood post.
2.4.3 From collaboration to grounding the conversation

We have seen that repairs can be made to correct an error, or to establish mutual understanding in a conversation. Building upon the collaborative model and the idea of mutual understanding is the concept of grounding. As this concept is prevalent in research about CMC and a continuation of research regarding collaboration in communication, it will be covered here. According to Clark and Brennan (1991), speakers have to find a common ground (mutual knowledge, beliefs and assumptions) in their conversation in order for it to be successful exchange of content and process. “All collective actions are build on common ground and its accumulation” (Clark, Brennan 1991:222), hence the name grounding. The purpose of grounding is defined in the grounding criterion: the contributor and his or her partners mutually believe that the partners have understood what the contributor meant to a criterion sufficient for current purposes (Clark, Brennan 1991:223) (See also: Clark & Schaefer 1989). This then allows grounding to be defined as the collective process by which the partners try to reach this mutual belief (Clark, Brennan, 1991:223). In a conversation words can take on a different meaning than expected for example. Two of the main factors of grounding are purpose, what participants are trying to accomplish in their communication and medium, the techniques available in the medium for accomplishing that purpose (Clark, Brennan, 1991:223). The purpose of a conversation is something that people try to establish collectively and can change in the course of the conversation. As the purpose of the conversation changes, so should the grounding techniques. When establishing what object is indicated, or what person is intended in the communication, a variety of techniques are available. These ensure that not only the listener is referred to the intended object, but also that the listener makes it clear to the speaker that he understood. The object is then part of the common ground.

Face-to-face communication, on which the above mentioned theories and principles are based, is of course only one of the communication media available. Different media have different grounding techniques and not all grounding techniques are (easily) available to every medium (Clark, Brennan 1991). Thus, it is possible to assign a “cost” to a certain grounding technique based on the difficulty of use of that technique given the medium in which it is used. Take the use of “okay”, which is easily inserted in face-to-face communication making it a low cost acknowledgement. But in online chat it is much harder to time correctly in the sequence, to type it without errors, and could interrupt the other participants in the conversation. The cost of the acknowledgement in this medium is therefore higher (Clark, Brennan 1991:229). Additionally, media can impose certain constraints on communication: audibility isn’t a concern for written text. A medium lacking a characteristic (of constraint) generally forces users of said medium to use
alternative grounding techniques. In order to minimize the collaborative effort needed for understanding, the messages in chat are longer and more carefully formulated in order to minimize the need for repair (Clark, Brennan 1991: 232). Finally there is an interaction between the purpose of the communication and the medium used for the communication. When it is necessary to have a written account of something like an appointment, email is a more preferable medium than face-to-face communication. When trying to reach a consensus, face-to-face communication is more preferable (Finholt, Sproull, Kiesler, 1990 via Clark, Brennan 1991:232). As an addition to the collaborative model of communication, the principle of grounding can influence the way redditors compose their message or manage their interaction in a thread. Redditors may have to make a choice between editing their own post or creating a reply to a post if they have to alter their message in order to ground the conversation.

2.4.4 Media richness theory

An often used theory to describe different media is the Media Richness Theory by Daft and Lengel (1984, 1986). It is possible to classify media in many different ways. A division between mass media on a one-to-many basis (TV, radio, newspaper and more personal media on a one-to-one basis such as telephone calls, or letters spring readily to mind. It is also possible to classify media according to the communicative possibilities a medium has, or how rich a medium is. Richness being the information-carrying capacity of data (Daft and Lengel, 1984). Rich media like face-to-face communication allow for a large amount of information to be sent, not only via speech but also by intonation, gestures, contact etcetera. Poor media, like spreadsheets provide little feedback, do not utilize many channels of communication (visual, audio etc), give little cues and cannot be personalised. The table below provides an overview of the richness of several media updated to include email and electronic data interchange (EDI).
Despite being criticised (see Dennis, Kinney 2000) in recent times the media richness theory allows us to make a distinction based on the amount of information available via a medium and the way in which this information can be presented. It provides a practical distinction allowing for a straightforward comparison between communication forms. Looking at the table and keeping in mind that Reddit only allows users to post text and links to images and/or video, Reddit is a moderate to poor medium that can allow for fast feedback, but by limited channels and can be both personal and impersonal in its communication. Combined with the aspects of costs, constraints and affordances mentioned above a cynical reader might ask why use Reddit or CMC at all or with such an intensity that it is one of the most visited sites on the internet?

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Media Characteristics That Determine Richness of Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium</td>
<td>Feedback</td>
</tr>
<tr>
<td>Face-to-face</td>
<td>Immediate</td>
</tr>
<tr>
<td>Telephone</td>
<td>Fast</td>
</tr>
<tr>
<td>Electronic (e-mail, EDI)</td>
<td>Fast</td>
</tr>
<tr>
<td>Written, Personal (letters, memos)</td>
<td>Slow</td>
</tr>
<tr>
<td>Written, Formal (bulletins, documents)</td>
<td>Very slow</td>
</tr>
<tr>
<td>Numeric, Formal (computer output)</td>
<td>Very slow</td>
</tr>
</tbody>
</table>

Note. Adapted from Daft and Lengel (1984).


2.4.5 Dealing with the limitations of CMC

Why indeed, if text-only CMC is apparently a poor, and claimed by some (Hale 1996), via Herring (1999)) incoherent medium is it so immensely popular? In Interactional coherence in CMC (1999) Susan Herring answered this question. Rather than taking incoherence as a determining factor for its use, i.e. the more incoherent a medium the less likely it is to be used, she states that the interactional coherence and its popularity are independent phenomena. Adding to that, she states that users may even be attracted to the incoherence of a medium and that users may even be put off by a medium in which coherence is strictly enforced. She states that text-only CMC can be both incoherent and enjoyable because of one of the key concepts of text- only CMC: a persistent textual record that can negate the perceived incoherence and which allows text-only CMC to move beyond an imitation of face-to-face interaction. We will cover this in more detail below.

Computers weren't originally intended for social interaction, and research into computer supported cooperative work showed that that CMC as a medium imposed several limitations on the
group interaction (Herring 1999:2). Specifically mentioned are the lack of simultaneous feedback and disrupted turn adjacency. The lack of simultaneous feedback, compared with face-to-face communication is due to the fact that in general messages cannot overlap and the reduced audio-visual cues (Herring 1999:2). Text-only CMC is, when held against Daft and Lengels media-richness theory a very poor medium. There is for example no way for a participant to see facial expressions, or to hear whether the intonation indicates a sincere or sarcastic reply. Additionally mult-participant text-only CMC systems (hereafter referred to as chat in this paper), with the most prominent example in the paper being Internet Relay Chat (IRC) generally allow only complete messages to be sent. As such, no participants other than the sender can intervene in a message, whereas interrupting someone mid speech is possible in face-to-face communication (Herring, 1999:3). This can mean that posts by participants may overlap and can make it difficult to provide a properly tailored message or reply (Herring 1999:3)

Disrupted turn adjacency means that messages are posted in the order they are received by the program, not as they are intended by the participants. Interventions from other participants than speaker or listener, or a delay between the sending and receiving of a message can change the order in which messages are intended to be read and the order in which they are received. In spoken interaction, conversation has a high degree of turn adjacency; relevant responses are spoken closely after an initiating turn. Adjacency pairs and sequences structure face-to-face interaction. Whereas on the other hand, disrupting the adjacency pairs may lead to fragmented interaction (Herring, 1999:3)

Compared to face-to-face conversation where intervention is possible, feedback is swift and rich, where relevant responses are closely grouped with initiating turns (for more on turn adjacency, see also Grice, 1975), follow each other in an orderly fashion, where interjections in an ongoing discussion generally are frowned upon and where conversation is deemed difficult to follow by its participants when these aspects are flaunted it may seem plausible that text-only, one-way (half duplex) CMC is so difficult in use that a coherent conversation is nigh impossible (Herring, 1999). However: “plausibility is not the same as inevitability” (Herring, 1999) and text-only CMC is very popular indeed.

The question then becomes: “how coherent are computer mediated interactions?” (Herring 1999:3) By providing an overview of research literature using conversation analysis Herring provides evidence of dysfunctional interaction management in CMC. There are two possible explanations for this enigma: turn-taking and sequential coherence On the one hand there is the aspect of turn taking. In face-to-face communication turn-taking ideally takes the shape of precisely alternating turns. The transition of turns from speaker to speaker is instant or with minimal delay,
there is no overlap and speakers take turns in an orderly fashion (Herring 1999). As anyone who has ever used any form of chat can attest: this is rare in CMC. Email can have a delay stretching to days and in chat, because users can only view a completed and sent message, can be constructing and sending messages to each-other at the same time, overlapping exchanges. Additionally, in group chat, other participants can intervene in an exchange by inserting unrelated messages.

CMC also does not exactly distinguishes itself in the orderly alternating of speaker turns. Initiations may simply be ignored by others, for example. When initiations fall on deaf ears, the initiator may simply send his message again or even spam/flood the channel until a response (positively or negatively follows) Especially in group chat as IRC responses from many participants can be aimed at one initiator (many-to-one) or in contrast be one message addressing more than one initiating message (one-to-many). The combination of inserting unrelated messages and not orderly alternating speaker turns makes that exchanges can freely intermingle in a medium as IRC. Herring provides examples where three different exchanges, or subtopics occur and overlap in a mere 15 lines of IRC-chat and posts in a listserv (a generic trademark for an email discussion list) (Herring 1999).

These difficulties in turn-taking and deviating from the face-to-face ideal may show that users have difficulty in determining the intentions of others in text-only CMC. It may be that a users initiative was lost in a heap of other posts, or that a reply is taking longer than expected, there could have been a technical error, etcetera. Users compensate by using redundancy: Herring shows that whilst 18% of single message posting participants were ignored, all participants posting more than two messages were responded to (Herring, Nix, 1997 via Herring 1999). Redundancy is used to compensate for the “noise” in chat and compared to a chaotic party where every conversation is held at once, at the same volume (Herring, 1999:7)

On the other hand there is the issue with sequential coherence, an umbrella term concerning the difference between face-to-face interaction where turns are expected to “belong together” and CMC where sequential coherence is not even the norm. What does this “belonging together” entails? In face-to-face interaction responses or follow ups to turns are expected to occur adjacent in time. (Herring 1999:7-8) From adjacency pairs or exchanges to sequences of turns making up (discourse) topics, turns adjacent to each other are expected to relate to each-other and confirm globally to the discourse topic (Herring 1999:8) structuring a conversation both temporally and subjectively. In short: following the cooperative principle of Grice (1975, see above). In chat, many intervening turns make sequential coherence the exception rather than the norm. In email, one message can contain a variety of moves combining what in face-to-face conversation would have been several turns. Summarised: physical adjacency does not mean pragmatic relevancy. Subject
coherence is also often violated in CMC. Herring (1997, via Herring 1999) found that a third of the messages in her samples violated the Gricean maxim of local relevance. On a somewhat larger scale it appears that users of CMC are also prone to drifting off topic during a (group) conversation. Entangling threads or shifting from one topic to the next one, only loosely or perhaps even only semantically related to the previous topic on a whim are commonplace in CMC. Of course, some amount of topic drift is also characteristic of group face-to-face conversation but is present in far greater extent in CMC. The not-following of sequential coherence, local relevance and topic drift is so much the norm that a bot (a program posing as a person in a chatroom/irc-channel) could fool a user, even when the bot's answers would make no sense whatsoever in a face-to-face situation (Turkle, 1995 via Herring 1999).

The popularity of CMC indicates that users have, at the very least, found workarounds for these problems or solved them outright. Herring states that on the one hand users adapt themselves and their actions to the medium they use, and on the other hand that the “limitations” needn't necessarily be limitations but can also be used as advantageous and helpful (Herring 1999).

Adapting to the limitations of text-only CMC enables users to deal with the absence of feedback present in face-to-face interaction. Users use alternative means of signalling connections across turns, negotiating turn alteration and signalling listership (Herring 1999). Where in face-to-face interaction backchannels, small responses made whilst listening to a speaker (Herring 1999, Wikipedia C) play an important role, they are deemed more rare in CMC (Herring 1999). Still, there are CMC environments where backchanneling plays such an important role that users have implemented “short cuts” to (quickly?) type them, providing feedback as “nods” and “giggles” and even a “questioning look”. As this is an example from a Multi User Dungeon (MUD, a text/chat based form of a multiplayer online role-playing game) these backchannels can perhaps be seen as responses in the chat. Another way to provide backchannels was found in a listserv, participants would quote a post they agreed with and provide their backchannels in the form of a response to the quoted text (Herring 1999:12).

Adapting doesn’t stop with backchannels. Reducing speaker change cost, or facilitating the timing of turn alteration is another area in which users of CMC have adapted. Users can use previously designated character such as % at the end of their message to indicate that although they have sent a line to other users, they in fact aren’t finished or ready to give up the floor (Herring 1999:12) Personally I have seen varieties of this where users would use an X/Y ie. 1/3 like indicator, telling others that their one message consists of three separate posts. In the role-playing setting of a MUD participants to the conversation had a command that would indicate that they'd be raising their hands (as in a classroom) and would be allowed to speak via moderator, creating a
classroom-like discussion (Herring 1999). In these ways, indicating who has, or doesn’t have the floor is easy in one-way CMC, but what about two way? Here users would adopt a overlap-pause way of typing, first filling the screen, and then reading and rereading until understood. Quite different from face-to-face communication where talking at the same time is hardly comprehensible.

We have seen that in many forms of CMC messages are placed underneath each-other chronologically. How then to refer to a specific messages some turns ago? In chat, the notion of addressivity is prominent: users preface their message with the (nick)name of their addressee. In other forms of CMC linking to previous messages or quoting them or bits of it are more prevalent. This way the illusion of proper face-to-face-like adjacency with all its benefits is created. Additionally, to avoid getting bogged up in a myriad of subjects and their corresponding conversations many forms of CMC are organised around central topics. IRC-chat has channels with relatively set subjects, as do Usenet newsgroups and email lists. To keep participants on topic there are moderators, who can steer a conversation (Herring 1999:14). In the case of IRC, moderators can “kick” misbehaving users out of the channel, or outright ban them from it.

Even with these adaptations incoherence appears to be the norm in CMC, this incoherence doesn’t appear to hinder its users, so might it be advantageous? Probably. Relaxing the norms allows for playful interaction, enabling collaborative joking sequences or the chaining together of jokes where the loose norms regarding relevance can be beneficial. By not strictly adhering to local relevance joke chains can switch subject far quicker than possible in face-to-face whilst still maintaining some coherence. Humour in CMC can be metalinguistic; making jokes on CMC about CMC. An example is given by Herring which doesn’t only concern endlessly embedding quotes in mail but also about its beginning message:

On 30 Jan 1998 00:42:15 GMT, http://root@tlerll.earthlink.net (Tlerll) wrote:

- In article <http://6ar3k6$a9c@newsops.execpc.com>,
- Frank Rizzo <http://dlouhy@execpc.com> wrote:
- >>> This means ABSOLUTELY NOTHING TO ANYONE EXCEPT ME.
- >>>then why the need to post
- >and why the need to complete sentenc
- *so why the need to verbs.

Besides cracking jokes CMC, and especially group CMC, offers the possibility of multiple
simultaneous exchanges. People can for example play simultaneously in different MUDs, playing a different character in each MUD. Generalizing; CMC enables participants to interact simultaneously with several groups of people at the same time, something physically impossible face-to-face. This form of communication exceeding face-to-face interaction is known as hyperpersonal communication, a term briefly touched by Herring, entailing the notion that: “CMC (can) surpass the level of affection and emotion of parallel FtF interaction” (Walther 1996, 17)

The term hyperpersonal warrants some more detail than given by Herring. According to Walther (1996:17): “hyperpersonal communication - (is) CMC that is more socially desirable than we tend to experience in parallel FtF interaction” and applies not only to social interaction but also when making decisions and business via CMC. Walther then mentions the hyperpersonal aspects of the senders and receivers, the characteristics of the channel and the feedback processes. It goes beyond the scope of the thesis to cover this in debt and we will limit ourselves here to a brief summary. According to Walther, the receivers in CMC can “inflate the perceptions they form about their partners”. Essentially; the fewer cues and less knowledge about a conversation partner, the more important whatever and how little cues and knowledge are received. Senders can optimise their self-presentation, self-censor and be very aware of how they want to be seen by others. Text-only CMC doesn’t allow many cues to be sent along with the message (it is in other words a poor medium), meaning that first impressions are more manageable than in face-to-face interaction: text-only means not having to worry about your appearance or other distractions or disabilities. Finally, the lack of physical interaction may mean that the sender can focus solely on the creating of his message. He doesn’t have to worry about sitting up straight or smiling at appropriate moments. (Walther 1996:16-22)). The hyperpersonal aspect of CMC is therefore twofold. Communication via CMC needn't limit itself to the physical limits imposed by the “real world” and CMC allows conversational partners to both mould and view a perfect image of themselves and others.

Conversational persistence, that is the availability of a persistent textual record of the CMC-interaction is what aids users in coping with the increase in incoherence. Conversations can be read and reread until they are made sense of. This textual record can be created by explicitly logging the conversations had in a program, or by its nature: email for example has to be deleted manually. Even if the conversations aren’t logged, users can still read and re-read whatever messages they have on their screen in front of them, enabling them to have multiple conversations simultaneously (Herring 1999) Lastly adaptation may come in the form of topical organisation: having different rooms, newsgroups, threads or channels based on their primary subject and with moderators enforcing that users stay on topic (Herring 1999).

What does these observations mean when they need to be applied in CMC design? Herring
names the following options available to improve coherence without sacrificing the attractive incoherence of CMC. Extensive and easily available logging of the messages, recorded in such a way that it allows users to have a clear view of the sequence in which the messages were received, is the first of the suggestions offered (Herring 1999). The second one is reduce the incoherence caused by insufficient feedback, switching from the prevalent one-way form of CMC to a two-way form, although this would cause logging problems and provide some challenges in realising two-way chat-like applications. Finally there is the reduction of incoherence by tracking and linking the logically connected turns. For example colour coding of the connected turns, fading away as they are no longer current. Or a completely separate screen displaying the logically coherent messages, something Herring envisioned as taking the form of topic chains or trees. With trees having the advantage that due to their nodes and leaves can incorporate hierarchical relations. Admitting that these solutions are problematic she leaves them as suggestions for future research (Herring 1999).

How do some examples of CMC other than Reddit cope with incoherence and how are the above suggestions implemented? The first thing to keep in mind is that 1999 is not only sixteen years ago, but in terms of technical development perhaps more akin to the computer stone-age. To provide some perspective: windows XP was released in 2001, Netscape was still a commonly used browser and iPhones were introduced in 2007. In the years after writing this there have been so many technological developments that it would be folly do describe them all. We'll limit ourselves to descriptions of some of the applications that tackled some of the incoherence of CMC as described before going into detail about the way Reddit deals with these problems. The coherence problems with CMC as described by Herring were a lack simultaneous feedback and disrupted turn adjacency. Users have adapted to these problems, but many forms of CMC have as well. Rather than a full blown adoption of two-way systems, one-way systems tend to be the norm. However, they have added several useful features to improve coherence. Applications like WhatsApp and Facebook Messenger feature a small message that the other person is “now typing”, giving participants the option to either continue or wait until the message is received. This simple addition can solve overlap and lessen the disrupting of turns. Herring mentioned the repeated sending of messages to get a response. Although applications as IRC don't mention when a message is received, applications that implement the “now typing” feature can prevent the repeated sending of a message. Participants using this application can of course still re-send a message if they want to, but are given feedback by the program that their message was received, or even read by the conversation partner. Less chatty, but still capable of a quick conversation are web forums. Released in 2000, phpBB allowed for the easy creation of internet forums allowing division not only by subject but also by thread. Take for example the phpBB forum of tech website arstechnica.com. It
has umbrella terms such as “hardware & tweaking” divided in several subfora leading to the individual threads filled with posts. As the forum grows, moderators keep users on topic and can hide or remove threads or allow them to remain as an easily searched archive.

To what extent and how does Reddit cope with these problems with incoherence? Reddit was founded six years after Herring’s paper was published, and as one of the founders was a computer science student, there is a chance that he read Herrings paper, although this is just speculation on my part. Still, it is most peculiar that Herrings suggestion of using a tree-structure for coherence is implemented almost ad verbum in Reddit. Albeit with letting users decide what essentially is the most coherent, highest scoring order is, rather than using an intricate computer system thus eliminating the need for very technical solutions. Redditors can reply not only to the first post in the thread, creating a new node, but also reply to a reply or even reply to a reply of a reply (etcetera), providing nodes that can become branches themselves. Using the tree-structure also allows a sense of order in a thread. One thread could have several conversations in them, all shown by the indentation of the posts and the lines connecting them. Still, compared to a phpBB forum I found Reddit to be more confusing at first sight. In my opinion Reddit has sacrificed visual coherence for textual coherence, and could profit by having a better visible distinction between posts in a thread. Although Reddit implements Herrings third suggestion for designing a CMC-system, I doubt she would be pleased (or is pleased) by its form: it is not two-way, does not cope with simultaneous posts and users can only post plain text, limiting direct feedback.

2.4.6 Repair in chat room interaction

Jacobs and Garcia's *Repair in chat room interaction* (2013) deals with the use of repair in a chat room. Like Herring above they state that chat can be read as a document (to start at the first message and working down to the latest), but adding to that they also state that chat can be read as a conversation (to start at the latest message and working up to the first) and that users of chat can primarily use either of these ways of reading or interchange them. By looking at data from small group chats, where not only the chat logs, but also videotapes of the participants screens were used, the researchers found not only the way the participants used repairs in chat, but were also able to explicitly note the difference with repairs in face-to-face conversation. Videotaping the screen allowed the researchers to see if participants for example paused or erased their messages. Due to the tempo of chat, participants may at times discard their message being written for a completely different one, editing their utterances in progress. They may do so because the conversation has moved so that their indented utterance is no longer relevant. As this is done in private, (others can't
see what they're typing, only the completed message) that type of repair isn’t visible for others. Repairs of that type are more error avoidance. Self repairs after a message has been posted may end up out of place as the conversation moves on and the repair is no longer adjacent to the error. Repairs initiated by others are more explicit in their wording than in face-to-face conversation, where a “huh” may suffice. This again is done as the conversation may have moved on and participants cannot be sure that their post is placed at the proper place in the conversation. Other-initiated repairs may not even be adjacency pairs, as would be the case in face-to-face conversation, as the participant needing to make the repair could be otherwise engaged and misses the post asking for a repair. As participants in chat rooms expect that this can happen it isn’t frowned upon or seen as improper conversation. The participants simply move on with the conversation. Its this difference in placement of messages in different places than expected or intended that differ from face-to-face conversation that has effect on the organisation of repairs (Jacobs, Garcia 2013: 565-578). Jacobs and Garcia found two types of placement problems: actual displacement, where posts of others are placed between the post and the post referred to and failure to anticipate potential displacement, where a participant fails to take potential displacement into account (Jacobs, Garcia 2013:578). To solve displacement problems, participant have a number of potential solutions. They can split or post shorter messages, delay their posts to ensure proper placing, use address terms to minimize confusion as to who is the intended reader or what is the intended post, using parts of a previous message to ensure proper pairing.

Reddit is not a chat program, and while it can have a high tempo of posts being placed, it is not fair to directly compare Reddit to chat. That does not mean that Reddit is free of displacement issues, or that we cannot make a valid comparison between the two forms of CMC. Beside the two types of displacement problems mentioned above, Reddit can have displacement due to the way it ranks responses. A message that solely contains a repair or a message that hopes to be the second part of an adjacency pair but when seen as being irrelevant or for any other reason is downvoted by the redditors, may become hidden from view. Whereas in chat messages can be re-read, even though they may require tentative amounts of scrolling, they are not hidden from view. Messages on Reddit scoring too low are hidden and require more clicks and more distinct action from readers. Reddit, due to its tree-structure, has much less actual placement problems than chat: as replies can be made more accurately. Rather than adding the messages to an indistinct line of posts, a reply is made to the appropriate node in the thread. By having a division in subreddits, threads and what boils down to branches in that thread replies can be made in a more specific manner. This is comparable to a party where small groups of people converse and a reply is made solely to a specific small group. Others chat as well, but do not intermingle with your current discussion. In
chat, this isn’t the case: the entire party converses at the same volume, and although messages may be intended for a sub-group, they are still sent to the entirety of the party. While this division of the thread doesn’t prevent other redditors from inserting their message, it does provide redditors with a smaller conversation window. That is to say, because the conversation is more structured, there are less possibilities for inserting messages that do not have anything to do with the conversation as going on in that particular branch of the thread. So, whilst the insertion of posts in between at the same spot is still a probability, redditors can cope with it more easily. Error avoidance, or repair in private is much the same for chat as for Reddit. The techniques for coping with placement problems in chat are found in Reddit, but to a lesser extent. Reddit uses some visual aides, the dotted lines and nesting to indicate the intended placement so that coping techniques as placing shorter messages aren’t needed. Research covering repairs in online forums wasn't found, or so psychological in nature and interwoven in the conversation that it couldn't be used in this research (eg. Kleinke, 2008).

2.4.7 Accountability

When analysing repairs in CMC we have to take accountability, “that is how, people make their messages achieve recognisable social and personal objectives while attending to the discursive perils attendant on any contribution to the social scene.” (Antaki et al, 2005: 114) into account. Or how “in any interaction, participants do things for which an account might, in principle, be called for and given; things that “make sense” interactionally and in the moral order” (Antaki et al., 2005: 114). It could be that the edit function of Reddit is used by redditors to achieve their objectives, to challenge another redditors accountability or to reformulate their post in such a way as to make more (accountable) “sense”. To study this to the degree of detail needed to provide well founded conclusions, would require in depth research we cannot provide with our research set-up and method. We agree with Antaki et al. (2005) in both their choice for a case-study based research, their conclusions and their remarks in their discussion section. They noted that users can exploit the difference between face-to-face interaction and the interaction on a public forum to create acceptable and accountable messages whilst breaking with the conventions of face-to-face communication. Secondly, their notion that although researchers can describe the actions of users of CMC in the most technical of terms this minimizes the story and applied knowledge of the participants is one we can only agree with (even though their criticism is based on questionnaire or interview based research). As we base our research on the reason for an edit given by a redditor, rather than having a redditor fill out a questionnaire, there is an aspect of the reason that can be lost
in transcoding. By looking at the edit or repair in the context of not only the post, but also of the thread it's posted in, we have minimised this. It is the inclusion of the context of a post in the research that is the most applicable lesson that can be applied in here. As this is more of a preliminary research into the use of repairs, we had to make a selection of criteria and research-angles that are investigated here. Accountability is a concept that, when combined with repairs and Reddit is something that warrants its own research paper.

2.5 Difficulties of using Reddit as a user and researcher

Reddit perhaps isn’t the easiest form of CMC to use. Without going into an usability-analysis of Reddit (a thesis on its own) we can point out some of the difficulties of using Reddit. Reddit is an online, text-based medium, a poor medium when compared to face-to-face communication. Redditors can post written messages and although these can contain links to images and video these aren't displayed (fully) on the website. Although other forms of CMC allow embedding of video's or voice-chat, Reddit does not.

When we make a quick comparison of Reddit with say, Facebook or Twitter, the primarily text-based interface stands out. Where Facebook and Twitter make ample use of buttons and icons as links in their interfaces, many of the links of Reddits interface are text only and aren't instantly recognisable as links. Reddit has a rather steep learning curve in my opinion. Being a poor medium, and text-based, all the constraints and limitations mentioned above apply to Reddit. These can make it difficult to proper grounding of the conversation, necessitating the need for repair.

One of the major constraints of Reddit is similar to other forms of text-based CMC; interactional coherence isn’t a given. When looking at CMC in the form of text-based chat, interactional coherence seems to be something that hardly is present, at least at first sight. The same goes for Reddit. Although the tree structure helps to makes sense of the wall of text thrown at a user, the tree-structure is perhaps hard to comprehend at first.

Then there are the difficulties of use of Reddit from a researcher's point of view. Obviously these difficulties needn't apply to other sorts of research, for example a usability-research wouldn’t necessarily require data from redditors, but these points need to be taken into account. The demographic of Reddit is unclear. While Duggan and Smith (2013) researched the demographic of Reddit users to find that a relatively small percentage of adults use Reddit and that the data is skewed to favour young males. This research was confined to the United States only; the conclusions drawn from that research may hold for the Western world, or for all (global) users of
the internet or simply just for the US. Research regarding the total demographic of Reddit wasn't found. This means that we cannot generalize the findings of our research to the entire internet demographic. A skewed demographic is not unique to Reddit. Twitter for example has a skewed demographic, at least for Dutch users (Tjong Kim Sang, 2012).

Perhaps one of the only things that can be said of all the users in this research with a reasonable degree of certainty is that they all understand or speak English. However, this does not even need to be the case as automated translators as Google Translate could facilitate non-English speaking participants here. Participants could come from any part of the globe and have a very different background when compared to other participants or the topic starter. This diversity only abets to the difficulties in grounding and creating a mutual understanding between participants.

2.5.1 Differences in repair use in Reddit and spoken interaction

Through primary research and observation of threads on Reddit for a longer time several differences between repair use in Reddit and face-to-face conversation were found. Not only in the way redditors use repairs, but also in the way repairs were and can be researched with Reddit. One of the primary differences in the research of Reddit and the research of for example Levelt is the fact that Levelt and others could observe far more of their subjects and their utterances when compared to this study. Where Levelt was able to record his subjects, in Reddit only the “final” edited item is visible to both visitors of the site and to the API to be mined. This may seem a trivial difference, but it could very well be that a large amount of repairs are made in the construction of the message, rather than after the message is posted although we can't be sure of that without further research.

Face-to-face communication is not permanent, unless recorded. Utterances are gone, once spoken. This naturally does not hold true for Reddit: written online communication has a distinct permanent character as long as the posts are not removed or deleted by either Reddit or the redditor that created the post. However, only the final version of a post is visible on Reddit. Unlike for example Wikipedia where changes to the lemmas are logged, Reddit simply shows the last version of the post and doesn’t archive the original posts. Thus, in Reddit, messages can be altered (edited, or deleted) without the original message being present in any shape or form afterwards.

Then there is the problem that adjacency pairs don't have to be adjacent to each other, as covered above, combined with the fact that the original, unrepaired post can't be recovered. And that Reddit does show that a post is edited, but that a reason for the edit doesn't have to be explicitly given by the poster, this means that an edited post can refer to another post quite a distance away in an ongoing discussion allowing even more noise to interfere with a correct
classification. As such we will have to classify every edited post that doesn’t explicitly state a reason for it being edited as a covert edit. Remember that the reddiquette states that a reason for the edit is preferable, but not obligatory. Edited posts where a reason is given within the post are thus seen as overt edits and can be classified further. Even when there is a reason given, in order to classify a post, the preceding and following posts still need to be analysed to be certain an edited post can be placed in the proper category. This as redditors can give a variety of reasons for their edit, and not all are clearly formulated and understandable.

2.5.2 Repairs and edits in CMC and Reddit

The research described above covers spoken interaction. While it is certainly possible to have a conversation via CMC, through Skype, or a VOIP application on a laptop or tablet, that is something not covered in this thesis. Reddit is a form of text-based CMC, and before going into detail we need to cover the way repairs can be made in several forms of CMC. The form covered in this thesis is the edit of ones own post, an in-post edit. CMC such as web forum also allow a user to edit their post. How a post can be edited on Reddit is shown above, and for webfora the concept is similar although the implementation is somewhat different. A user, or in some cases a moderator can make changes to the post after it has been placed, and the post stays on the same position in the conversation. Forms of CMC that don't group the messages, such as e-mail can't implement such a feature; it is up to the sender to send a new message repairing a previous one. This also gives e-mail a far more permanent character than Reddit or forums: repairing as a mechanism is possible, but the message cannot be altered any more.

Another way repairs can be made in CMC is by use of a repair marker, like the * (asterisk) in the following example, although other characters can be used as markers. Especially in online chat, where the original message is sent to the readers in its entirety another form to indicate a repair is used. Here the message gets corrected, immediately or quickly after detection of the error by a new message containing an asterisk (either at the end or beginning of the repair). If for example a message contains a spelling error or typo the next message may correct that. *for. This type of repair was not commonly encountered on Reddit during the preliminary research. Asterisks were more commonly found as a way of creating a bullet point like list, or to indicate some form of movement: “*bounce * *bounce * *bounce *”. Only a few examples of the use as asterisk as an indicator of an edit were found like the spelling correction in the third line:

1 Fellow Torontonian here. We brave through are shitty winters thank you very much
Winnipegger here. I laugh at your winters.

Both a slash (/) and asterisks are used to indicate that in the first post contains a spelling error, “are” rather than “our”. Although redditors don’t commonly use a repair marker like an asterisk, using an asterisk in this manner is apparently so popular in other forms of CMC that Reddit indicates an edited post with an asterisk.

One of the reasons I assume that this type of repair isn’t commonly found in Reddit is the way the posts are displayed by default. By default posts are sorted by the score they have. A post with a score that is too low will be hidden from sight. Also, it is easy for posts to be separated by many interceding posts. In this example there is already a post between the reparandum and repair.

Another difference that has to do with the visibility of the subjects for the researcher is that covert repairs can take different shapes in CMC. Smith (2009) notes that hidden repairs in a synchronous computer-mediated communication environment can take the form of scrolling and points out the shortcomings in merely transcribing the computer mediated communication, rather than using screen capture software in order to gather more analysable data. When mining Reddit as later described in this thesis, seeing the creation of a post isn’t possible and as such, a large amount of the edits that are made in the creation of a post aren’t seen. This doesn’t mean however, that edits of posts cannot be classified. Analysing repairs whilst a post is being made is perhaps something for future research.

As Reddit is a written medium, all the differences between spoken and written interaction apply. One of the most prevalent is the fact that spoken interaction isn’t permanent, whereas the messages on Reddit are easily re-readable. This has consequences for the interaction and the uses of repairs. Although it isn’t hard to imagine a situation in which a spoken utterance is repaired several times, utterances such as: “could you grab me the jar from the shelf, no the red one, no the other red one, on your left, your other left, no wait, I don’t mean red, it's more a dark brown, yeah, that’s the one!”, examples from Levelt (1983) generally contain one, perhaps two edits per sentence or turn. Clark and Brennan (1991) noted that written conversation was far more explicit in nature and shorter in form than face-to-face conversation.

Preliminary research showed that many posts containing repairs contained more than one repair. One of the recurring statements made in the research of repairs in spoken interaction was that repairs impair the flow of conversation (Plug, 2011), (Postma, Kolk, 1990), (Schegloff, 2013),

Source: https://www.reddit.com/r/IAmA/comments/2l1e78/iama_resident_of_a_small_town_in_hawaii_where_an/?limit=500
(Levelt, 1983). In Reddit, it appears that the repair does not necessarily interrupt the conversation. The preliminary research appears to suggest that using edits in fact aids the conversation. Take the following post for example. It is taken from the IamA-subreddit, where people with interesting experiences or professions answer questions from others:

*I have lived in Pahoa, Hawaii, for the past 24 years. I graduated from Pahoa high school and now work there. [The Hawaii Volcano Observatory site](http://hvo.wr.usgs.gov/activity/kilaueestatus.php) and the [local county civil defense updates](http://www.hawaiicounty.gov/active-alerts/) are my go-to sites but I drive past the flow on my daily commute. A lot of lives will be changed by this and it is some interesting times out here. The national news has been putting out some misinformation and I'd like to give folks the opportunity to ask questions about it.*

*Edit: Not sure what to provide as proof but [here's my driver's license](http://i.imgur.com/LAhhkut.jpg)*

*Another Edit: [Good site with up to date news and videos](http://www.bigislandvideonews.com/2014/11/01/video-morning-lava-flow-update-saturday-nov-1/)*

*Another other edit: I will be back on later tonight to answer more questions! Thanks for helping this situation be heard Reddit!*

*Super neato bonus edit: Holy spam, front page! I'll be back in the morning to get back to as many of you as possible! I'm so glad people find this interesting! CNN, get yo' facts straight!*

*Obligatory front page edit: Thank you SO much everyone! I am still here and trying to get to questions! I want to mention for those who are asking about helping - The Red Cross is here helping if you want to donate to them, but I personally want to plug Aloha Ilio! Aloha Ilio is a network of volunteers who foster abandoned/unwanted dogs and find forever homes for them. As a result of the flow many people are moving and unable to keep their pets and abandoning them. Aloha Ilio volunteers take dogs in to their homes and commit to keeping them until they can house them*
with a new owner. Please consider donating to them or adopting a doggie! [Here is their site](http://www.alohaiiorescue.com/adoptable-dogs) but they are very active on facebook.

_Mahalo again so much for reading! We received very little attention after Hurricane Iselle earlier this year and no FEMA funding for a devastated community that was out of power for weeks. It is important that people know what is happening and I thank you for your interest and aloha!_

In this example there are five overt edits, and perhaps one covert edit “Mahalo again […] aloha!”. The overt edits all provide an extra to the ongoing conversation. The first edit, regarding “proof” is about the fact that the subreddit has a rule requiring some form of proof that the person is in fact who he claims to be. Although it is just an assumption, it could be that he was contacted by the moderators of the subreddit to provide some sort of proof, else his thread would be deleted. This could be seen as an interruption of the flow of conversation between the moderators and the poster, but given that that is a different conversation it does not interrupt the conversation between the poster and other redditors. There are two edits about being on the front page of Reddit, with one of them providing background information, something also done by another edit. Using edits in this way, or for example to provide updates on an ongoing situation could actually help the conversation as all the relevant information is available in one single post that either through the scoring system or the fact that it is the first post of a thread is highly visible at the top (or the actual beginning) of a thread. It appears that brevity is key on Reddit.

### 2.6 Additions to the classification of Levelt

Brennan quotes Schneiderman (1986, via Brennan, 1998) “… _human–human interaction is not necessarily an appropriate model for human operation of computers.[…]_”. Strictly speaking Schneiderman and Brennan are mentioning the interaction of a human with a computer-interface or program, specifically the difference in difficulty interacting with a system using a command line interface or a graphical user interface. Still, in order for a person to interact with another (group of) persons via a computer, the same limitations apply, especially when it comes to grounding the conversation. Therefore solely using a classification system intended for face-to-face interaction as Levelt’s would be incomplete or even inappropriate. Levelt’s system is still a good starting point for
classification of interaction but several categories specific to human computer interaction and Reddit have been added. This because users can have reasons for editing their post that are unique for Reddit. Already in the preliminary research other reasons for editing a post other than those presented by Levelt became apparent. Posters might edit their post for the reasons given by Levelt but also edit their posts for the following reasons:

The first reason is for clarification: expanding on ambiguity reductions, redditors can use the edit function to either give or ask for clarification. Rather than writing a new post, the edit function is used. Although classifying clarifications like these as ambiguity reductions would be possible, clarifications like these apply to a rather large part of a previous post, or even the complete post whereas Levelt only focuses on a word within an utterance.

You could overwrite it 10 times and using special tools you could potentially recover overwritten data, the only way to be completely sure you destroyed the data is destroying HDD itself.
That is why all the bigger companies do exactly that, if there was any other way I am sure they would know, yet they still do it the old school way.

**EDIT:** I said "the only way to be **completely sure** you destroyed the data is destroying HDD itself."
And for those uninformed people that think it is as easy to get rid of any traces on hdd as completely rewriting HDD, keep downvoting me I hope to reach -100 soon.³

Above is an example of an ambiguity reduction, emphasising the word completely. When looking at a clarification edit we see a more elaborate explanation spanning over several sentences as demonstrated in the excerpt from a post below. Here the edit clarifies a particular aspect of the Israeli/Palestinian conflict

I was in Israel and Palestine for a few months in 2010, working with multiple NGO's throughout the West Bank and Israel. Most of my time was spent traveling the West Bank extensively and studying the conflict, as it has been coined. By no means am I an expert or anything of the like, just saying I've had some experience. Below I'm referring to the situation as I understand it, that has been inherited by both the Israelis and Palestinians today. [...] ³

³ Source: http://www.reddit.com/r/explainlikeimfive/comments/2l2jpd/eli5_why_does_it_take_hours_to_install_big/
**Edit 2:** Because everyone is focused on the "apartheid", I would like to clarify. The majority of Palestinians I spoke with view their situation as apartheid...whether it is or not is up for argument. I wouldn't know exactly what to call it. The situation has many similarities to situations throughout history, but is in itself very unique. Their interpretation of their situation is that it is, apartheid. That is the word they spoke to me.\(^4\)

Redditors were often seen editing their post with some sort of thank you notice when the post was given a large number of upvotes by other redditors. A similar sort of edit was found when a post received a large number of downvotes that the author didn’t agree with, condemning the redditors who were downvoting the post. Edits of this sort were classified as This is classified as **Thanking/condemning**. Although not unique for Reddit, as for example image-sharing site Imgur has a similar system, the upvote-downvote system and subsequent edits are prominent enough for Reddit to warrant a separate classification, as in Levelt’s system these edits would merely be grouped under “rest”. Also falling under this category are edits thanking for other reasons, comparable to a so called **shout-out**; “an expression of greeting or praise that is given to someone in the presence of many people” (Miriam-Webster, 2015).

Redditors can also use the edit function for **Providing background**: by giving links to background information, or by providing an explanation for some phenomena. This is far more easily and less intrusively done than in spoken interaction: a hyperlink can be added, and adding several sentences in the post is not seen to break the conversation by redditors.

**Edit:** [Secure Deletion of Data From Magnetic And Solid State Memory](http://www.softpres.org/cache/SecureDeletionOfDataFromMagneticAndSolidStateMemory.pdf)

Just spent 20 minutes reading this article. I’m not surprised about efficacy of recovering data, but I did remember about failsafes of writing bits next to each other. The article goes through several iterations of "erasing" data, or degausing.

So I guess if one can get the difference between the bits of old/new it can be recovered, if they can’t, then it doesn’t sound like it. But the article is also from

---

\(^4\) Source: https://www.reddit.com/r/explainlikeimfive/comments/2anwqq/eli5_israelipalestinian_conflict_gaza_july_2014/
1996, I don't feel like researching more in detail at this point.

-------------

It was in a similar thread a couple of months ago, and I can't find it in my bookmarks. It was published from what I think was a University professor around 1999.

Either way, it's simple, the proof of burden is just to show that something is recoverable from a completely over/rewritten drive.

And the general consensus from whom? I have recovered many filesystems and files in the past. Some of which I expected was overwritten but probably wasn't. It's really not an easy google search to find something other than yahoo answers for an academic question, especially when "specialized hardware" is required.5

Continuing the conversation was also found: whether done for brevity, or taking advantage of the points a post already has, continuation of the conversation via an edit rather than making a new post is another distinct use of edits in Reddit. Reasons for doing this appear to be diverse, the edit below for example could be seen as a conversation ender and thus could have been be placed in a separate post, but was instead placed as an edit.

That's a really creepy thing to say anywhere, but saying it in here makes it a million times worse.

edit: well now that you edited it to make it like a billion times worse I have to say you are a really shitty person. I hope your alarm doesn't work next morning and you wake up late and try to run to work and hit something very hard with your little toe.

Continuing the conversation could also take the shape of questions that came up after other redditors provided initial replies.

It can happen that a post contains a statement, a question etcetera that turned out to be answered in some external source or elsewhere in the thread. If a redditor sees this he can use a repair to correct the mistake made in his post (Retracting a statement), providing other readers with

5 Source: http://www.reddit.com/r/explainlikeimfive/comments/2l2jpd/eli5_why_does_it_take_hours_to_install_big/
correct information. Edits of this kind can take the shape of the example below: a simple variation of “nevermind” followed by the reason for the edit.

*I thought she was going to extend it?*

*Damn, this sucks though. When I first read the article I got chills. I still cannot fathom the idea of this, and how it works. I really hope more people can end their life more peacefully, the way she did*

*EDIT: nvm, she says in the article she had a seizure the day after the Grand Canyon.*

Although Levelt’s category of phonological errors cannot be replicated in a written medium as Reddit, classifying spelling errors as a separate category closely related to phonological errors rather than lexical errors seems to make sense. Unlike lexical errors, where an entire word is used erroneously, a spelling error can be easily made within the intended word. However, spelling errors were classified as lexical errors for the following reason: unless a researcher was present during the writing of the post by the redditor, the severity of the lexical error cannot be determined. It could be a simple misspelling, or (especially in the case of commenting via a smartphone or tablet with autocorrect) a completely different word. By perhaps erring on the side of caution, spelling errors were grouped with lexical errors.

If there is a post containing multiple edits, these edits may be grouped together and seen as a separate “multiple edits” repair. The preliminary data had several examples, and multiple edits in a single post were also found during the observatory phase of the research. However, doing this might be close to an unsatisfactory large “rest” category. A post containing a background information edit, combined with a retracting edit likely has a different impact on a post with several thanking edits. There is the aspect of information-loss to consider: multiple edits may have a single purpose, and the post could thus be categorised as belonging to a single category or the all of the separate edits perform different functions and need to be categorised as such. In general, I believe the last option to retain as much information as possible regarding the separate edits, and will use the last option for categorisation of the data.

However, an exception to this rule must be made for the following edited posts: thread-starting posts or posts so highly rated and so often edited that the edits become an ongoing story. As these tend to take the same shape as live-blog reports (as seen for example in ongoing sports events)
these posts can be grouped as a single category: live-blogs or live-blogging. Whenever a post with more than one edit in this form was found, it was categorised as live-blogging.

In Conversational Repair and Human Understanding (2013), Schegloff mentions several operations, or actions that a repair might have within a conversation and places those in categories. Many of the categories he mentioned require that the original text is present to examine the intended operation. Schegloff mentions replacing for example: yet without knowing what is replaced, using replacing

---

https://www.reddit.com/r/worldnews/comments/2rmace/shots_fired_at_french_magazine_hq/cnh6mkn
as a category is nigh impossible. As mentioned above: in Reddit and the Reddit API it is impossible to see exactly what was edited in a post. For this reason only edits conforming to the reddiquette were used. Schegloff gives ten examples of operations of repair: replacing, inserting, deleting, searching, parenthesizing, aborting, sequence-jumping, recycling, reformatting, and reordering. Inserting means that extra words are inserted in the sentence, or turn. In deleting words are left out. Searching is best described as the act of thinking out loud. Parenthesizing is the inserting of a turn construction unit in an already ongoing turn, and returning to the original subject afterwards. In aborting, a turn is prematurely abandoned, either in favour of a completely new one, or to revise the turn to better fit the intended meaning. Sequence-jumping is the use of repair to make a quick change of subject possible. In recycling, part of the just uttered TCU is re-used for example to repair overlap with a previous turn, in reformatting the repair changes the grammatical construction of the TCU and in reordering the words in the TCU get re-ordered to work out the way the array of elements in the turn should look (Schegloff, 2013).

All these operations are based on face-to-face interaction and even though these repairs are self initiated self repairs, contained in a single turn, their categorisation depends heavily on the context in which they are placed. The categories of Schegloff have some similarities with the categories specific for Reddit and with the repair categories of Levelt. Reformatting is not more than the correcting of a grammatical error. Parenthesizing and sequence-jumping have some overlap with continuing the conversation. All Schegloffs categories are very focused on spoken interaction. Searching for a word is a normal (at least in this researchers opinion) thing to do in spoken interaction, but seems somewhat preposterous or overdone in text-based CMC. Recycling a part of an utterance to solve a problem with overlapping turns has no use in Reddit, where although there can be overlap in completed posts, the overlap in posts during their creation is not visible for others and thus cannot be cause of any problems necessitating that form of repair. Whilst we by no means mean to discredit Schegloffs findings, the operations he describes are either not applicable to text-based CMC such as Reddit, or have an overlap with the already used categories.

However an interesting remark made by Schegloff (2013) is regarding the use of the word repair, and the difference in vernacular English and the use of the term repair in communication science. An analogy between a tailor either repairing a torn jacket, or suggesting to tailor the jacket so that it fits better allows Schegloff to introduce the term “altering” for repairs that do not necessarily repair an error but rather provide a better or clearer way to formulate an already understandably (and effectively not in need of repair) message (Schegloff, 2013: 46-47). Several of the additions to the categories of Levelt can be seen as such alterations. Providing background for example is not making the text in the post more readable, but can provide an easy way for others to
better understand the post or the motives of the poster.

The categories mentioned above (clarification, thanking, providing background, continuing the conversation, retracting a statement, live-blogging) were added to Levelt's categories. The table below provides an easy overview to the categories that will be used to categorize the data along with the subreddits used in the gathering of the data. We will cover the choice for the selection of the subreddits in paragraph 3.1.2

Empty table containing the categories of Levelt with the added categories based on observation and preliminary research of Reddit.

<table>
<thead>
<tr>
<th>Subreddit:</th>
<th>Worldnews</th>
<th>Iama</th>
<th>Movies</th>
<th>News</th>
<th>ELI5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covert</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambiguity red.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambiguous coherent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexical error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syntactic error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thanking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Providing background</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuing the conversation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retracting a statement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Live-blogging</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

By adhering to the maxims of Grice, or intentionally not adhering to them, the participants in a
conversations know what to expect of each other in the conversation. By working together, the participants actively try to create a common ground between them and try to achieve this common ground in the most efficient way possible for both parties. In many of the examples given in the papers mentioned above the common ground is established without effort and the goals achieved in a quick and flawless way. However, what to do when speaker and listener cannot establish a common ground on which to communicate, or perhaps on a more basic level, simply cannot hear each other correctly? The answer to those questions is beautiful in its simplicity: they need to repair their utterances. Repairing an utterance, or editing a post, is done in a quick and easy manner on Reddit. Although the differences between Reddit and face-to-face communication are plenty we are still able to use Levelt's categorization of repairs as a starting point for our research. Preliminary research and observation of Reddit showed us that we have to add several categories to Levelt's in order to minimize the rest category.

3 Research method

In this part we will set out our research method. To find out how redditors use repairs we must first gather a corpus of repairs. In paragraph 3.1.1 we cover the data mining of Reddit through Reddit's API and the PRAW-wrapper (Python Reddit API Wrapper), describing how the Python script used for mining works and the pitfalls encountered during the writing of the script. This will be followed by explaining the choice of the subreddits used for the mining of the data and the extraction of the repairs from the datasets.

3.1 Mining Reddit

Reddit provides us with a gargantuan amount of data. Using Reddit's own API and a Python wrapper for interacting with the API, several subreddits were mined for data. After some initial setbacks preliminary research showed that edits could be extracted from the datasets using some of the standard tools available in Ubuntu Linux.

3.1.1 Using the API
In order to get the raw data we have used Reddit's API. This API gives us easy access to Reddit's data, provided we followed the rules set out by Reddit. There is, for example, a limit to the amount of requests per minute and by logging in with a Reddit user name one is able to send more requests per minute. Breaking the rules could result in the program being forbidden to access the API (Reddit e, 2015). To query the API without breaking the rules we used a pre-existing wrapper. This wrapper, PRAW\textsuperscript{7} complies with all of the rules set out by Reddit. PRAW uses the Python scripting language, and although there are wrappers designed in other programming languages, PRAW is arguably one of the most complete and well documented wrappers for Reddit.

Reddit, its API, the rules for using the API and PRAW are constantly in development. This caused some problems with PRAW not being able to query the API correctly and certain lines of code becoming deprecated during the writing of the Python script. Some solutions to these problems were found on the \texttt{/r/redditdev} subreddit, others via the documentation of PRAW and the JSON page of Reddit on Github. Other problems with Python were solved via the IRC-channel of Python. Although all the problems were solvable and while frustrating at times did not terribly hinder the completion of both the script and the mining of the data, it showed that being dependable on third parties is not preferable. To minimise being dependent on ever changing third parties, we tried to keep the gathering and processing of data as local as possible. We have only used PRAW to mine complete threads from Reddit. All other data processing was done locally and in separate scripts. An added benefit of this method is that it creates static datasets to work with. Reddit obviously is not a static website: new threads or posts get added, changed and possibly even deleted. The data is therefore a representation of the thread at the moment of mining, changes made later on in the thread aren’t available.

During the writing of the script, we ran many test rounds primarily on the subreddit especially designed for this purpose (\texttt{/r/test}) and on the Python subreddit (\texttt{/r/python}). After some corrections to the code these subreddits were mined without problems, however when attempting to mine a completely different subreddit (\texttt{/r/france}) the script crashed without exception. The solution to this problem lay in the different character-encodings used by Python and Reddit. By default Python uses ASCII for character-encoding, but Reddit uses UTF-8. As UTF-8 allows for the use of more characters than ASCII, the script encountered input it couldn’t recognise and returned an error. As ASCII is an integral subset of UTF-8, the script worked fine on subreddits that only contain ASCII-characters. Both \texttt{/r/test} and \texttt{/r/python} didn’t contain UTF-8 characters outside of the ASCII subset while testing, but \texttt{/r/france} did. \texttt{/r/france} is a subreddit primarily in French, a language speckled with accented letters not found in English, or ASCII. Explicitly stating the use of UTF-8

\footnote{https://praw.readthedocs.org/en/v2.1.16/index.html#}

55
encoding solved this problem

A problem of the script, that may need to be addressed in the future is that all the posts in a thread are collapsed into a single list. Replies can be made to the top post, but also to other posts in the thread. Although this doesn't result in unusable data, analysing the context of the repairs and the place in the discussion within a thread is made more difficult than it could be.

3.1.2 Choosing the subreddits

Having tested the script, it’s time to gather a large and diverse enough dataset, albeit one that could still be analysed manually. Rather than randomly grabbing data, or grabbing the most recent data, we chose to gather data from several subreddits in a way that provided a snapshot to an ongoing discussion.

Reddit has a large amount of subreddits, some more active than others. During the testing of the script we used several randomly chosen ones, including the subreddits mentioned above. Whilst usable for testing purposes, these weren't suitable for the making of our datasets. To get our datasets we selected several popular subreddits. Aiding in the selection of the subreddits were the popular suggested subreddits displayed in the grey bar at the top of the front-page.

On the front page of Reddit there aren't just the posts that “made the front page” (are so interesting or popular that they are featured on the front page), but also a large selection of subreddits that may interest the viewer. These are displayed in a bar at the top of the website. This selection, made by Reddit, was used to choose the subreddits from which to mine the raw data.

There are other ways of selecting the subreddits, yet we found these to be unsuitable for our purpose, we will cover them here and explain our selection method. The front page provides us with
a large, random-like selection of posts and their corresponding threads. Although these posts that made the front-page could have been used, they were not, as in most cases they did not contain enough replies or repairs to be suitable for mining. We assumed that a large thread would contain more examples of edits than smaller threads and confirmed this with preliminary test runs. As such we shifted our focus to the “hot” section of several popular subreddits (/r/news, /r/worldnews, /r/explainlikeimfive, /r/iama and /r/movies). The “hot” section of a subreddit does not contain the newest threads, but rather the threads that are popular within that subreddit at the time of the mining. This means that the threads found in the hot section of a subreddit in general have a large amount of replies and thus are suitable to be used as raw data. Additionally, these subreddits tend to have a conversation-like structure unlike other subreddits where we found the replies to be a mere collection of replies. An example of such a a popular subreddit with many replies, but little discussion is the subreddit /r/jokes (seen in picture 8). The picture shows four visible replies to the first joke. Only one of the replies is a reply to another reply. The other posts are in reaction to the starting post. Consequently: /r/jokes and other subreddits that exhibited the same pattern of replies were not considered for mining.

The subreddits /r/news, /r/worldnews, /r/explainlikeimfive, /r/iama, and /r/movies not only have a large amount of replies but also have a different reply-pattern than /r/jokes. Replies in these threads are not only made to the initial post, but also to other replies and take a more general discussion-like shape. The subjects these subreddits cover also lend themselves for debate. /r/news is a subreddit that covers (mainly) news originating from the US. /r/worldnews covers all news but the news originating from the US. During observation, these subreddits were found to contain links to controversial news messages, sparking debates. /r/explainlikeimfive, explain like I am five is a subreddit that allows redditors to have difficult questions explained to them in a simple and understandable manner. Thread on /r/explainlikeimfive were found to not only contain answers to the question asked but also follow-up debates regarding the validity of the answer, or a variety of other subjects. /r/Iama (ask me everything) is an interview-like subreddit where people who do something uncommon or who participated in an interesting and unique event (Reddit f, 2015) can be asked a wide variety of questions. The discussions that follow appear to be of a more personal type. /r/movies was chosen as it is a subreddit that concerns itself with movies, something one can like or dislike, providing discussion and incidentally proving the dutch proverb: “over smaak valt niet te twisten” (one cannot argue about taste) wrong.
The subreddits /r/news, /r/worldnews, /r/explainlikeimfive, /r/ama and /r/movies were thus initially chosen for their popularity, the large number of replies and the relatively large number of repairs in their threads. Additionally, the replies in the chosen subreddits were seen to take the shape of either a conversation or a debate between several participants, closely resembling a conversation one might have in a face-to-face conversation.

3.2 Extracting the repairs from the raw data
In order to minimise dependency on third parties the Python script was only used to gather threads from Reddit in its entirety, which were then saved and processed locally. From the hot section of the subreddits the top 10 threads were mined, labelled as raw data and saved as five separate files. These files contain the complete threads as seen on Reddit, combined with a selection of the available metadata. Each post is placed on a separate line, with a timestamp set by Reddit at the front. This timestamp is set in UTC-epoch seconds (*Universal Time Coordinated* or *Universel Temps Coordonné*). New threads are indicated with the addition “[new post]” and a timestamp, replies are also preceded by a timestamp in UTC-epoch seconds. These timestamps were added to have the date of creation of the post embedded in the file. Additionally, the filenames contained the day they were created.

Having gathered the datasets the posts containing the edits need to be indicated. As mentioned above, an edited post is given an indicator (“*”’) to show to other users that an edit has been performed on the post. This asterisk isn’t visible in the threads mined via the API. There is however a variable `Submission.edited()` in the metadata of a post. By default this variable is set to false. As an edit is made, this variable is set to a timestamp in UTC-epoch seconds. Although future research may use the timestamp to determine, for example, when an edit took place in the timeline of the thread and its effect on the conversation, the timestamp is only used here to give an objective indicator of an edit in a post. After all, a redditor may have written something like: “I am an editor for my school paper, these are the things I edit:” creating a false positive if the search was only done on posts containing “edit: “. As the script runs and finds edited posts it marks them with an “EDIT-EDIT-EDIT” line in front and an “END OF EDIT-EDIT-EDIT” line at the end of the edited post. Combined with the adding of the [new post] tag described above, these are the only change made in the data mined from Reddit. The posts themselves are taken as is, including spelling errors, adult language etcetera. This raw data (43308 lines in total) was then refined, as we are only interested in edited posts and looking through the complete threads by hand searching for edited posts is too time consuming. By using readily available tools in (Ubuntu 14.04) Linux such as *grep* we extracted both a list with all the edited posts, and with the edited posts were a reason according to the reddiquette was given.

---

4 Results

After gathering, extracting and refining the data we have found the following results and grouped them together in tables. Rather than opting for one large table containing all the results we split our findings over several tables to improve readability. In the table, there are instances of 0, meaning that a repair was not found in that particular subreddit but also a X in the case of the “different” category. This was done as it is impossible to see what the message was prior to the repair.

Table 2: Percentages of repairs

<table>
<thead>
<tr>
<th>Type of repair</th>
<th>Worldnews</th>
<th>Iama</th>
<th>Movies</th>
<th>News</th>
<th>ELI5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covert</td>
<td>0.81</td>
<td>0.61</td>
<td>0.65</td>
<td>0.74</td>
<td>0.45</td>
<td>0.72</td>
</tr>
<tr>
<td>Overt</td>
<td>0.19</td>
<td>0.39</td>
<td>0.35</td>
<td>0.26</td>
<td>0.55</td>
<td>0.28</td>
</tr>
<tr>
<td>Different</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Ambiguity reduction</td>
<td>0.00</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>0.05</td>
<td>0.09</td>
</tr>
<tr>
<td>Appropriate level</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Error (uncategorisable)</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.09</td>
<td>0.03</td>
</tr>
<tr>
<td>Lexical error</td>
<td>0.05</td>
<td>0.16</td>
<td>0.00</td>
<td>0.11</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Syntactic error</td>
<td>0.02</td>
<td>0.06</td>
<td>0.00</td>
<td>0.02</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>Clarification</td>
<td>0.15</td>
<td>0.02</td>
<td>0.10</td>
<td>0.02</td>
<td>0.23</td>
<td>0.09</td>
</tr>
<tr>
<td>Thanking/Condemning</td>
<td>0.09</td>
<td>0.22</td>
<td>0.20</td>
<td>0.14</td>
<td>0.05</td>
<td>0.14</td>
</tr>
<tr>
<td>Providing background</td>
<td>0.27</td>
<td>0.27</td>
<td>0.35</td>
<td>0.34</td>
<td>0.14</td>
<td>0.28</td>
</tr>
<tr>
<td>Continuing the conversation</td>
<td>0.27</td>
<td>0.18</td>
<td>0.20</td>
<td>0.18</td>
<td>0.27</td>
<td>0.22</td>
</tr>
<tr>
<td>Retracting a statement</td>
<td>0.04</td>
<td>0.02</td>
<td>0.00</td>
<td>0.05</td>
<td>0.09</td>
<td>0.04</td>
</tr>
<tr>
<td>Live-blogging</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Rest</td>
<td>0.07</td>
<td>0.08</td>
<td>0.05</td>
<td>0.09</td>
<td>0.00</td>
<td>0.07</td>
</tr>
</tbody>
</table>

This table contains the percentages of the repairs.

Table 1: Tally of the repairs

<table>
<thead>
<tr>
<th>Type of repair</th>
<th>Worldnews</th>
<th>Iama</th>
<th>Movies</th>
<th>News</th>
<th>ELI5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covert</td>
<td>235</td>
<td>79</td>
<td>37</td>
<td>126</td>
<td>18</td>
<td>495</td>
</tr>
<tr>
<td>Overt</td>
<td>55</td>
<td>51</td>
<td>20</td>
<td>44</td>
<td>22</td>
<td>192</td>
</tr>
<tr>
<td>Different</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Ambiguity reduction</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Appropriate level</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Error (uncategorisable)</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Lexical error</td>
<td>3</td>
<td>8</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>Syntactic error</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Clarification</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>Thanking/Condemning</td>
<td>5</td>
<td>11</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>Providing background</td>
<td>15</td>
<td>14</td>
<td>7</td>
<td>15</td>
<td>3</td>
<td>54</td>
</tr>
<tr>
<td>Continuing the conversation</td>
<td>15</td>
<td>9</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>42</td>
</tr>
<tr>
<td>Retracting a statement</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Live-blogging</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rest</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>13</td>
</tr>
</tbody>
</table>

Totals of the no. of repairs found for each category
The amount of edits confirming to the reddiquette is remarkably low. Of a total of 687, 192 (27.94% of the total) repairs provided a reason using the preferred “edit: “ prefix. Although this does not necessarily mean that the other 72.05% of the repairs provided no reason in their edited post, they did not provide a reason following the “edit:” guideline set out by Reddit. As such, they were unclassifiable and categorised as covert repairs. The overt repairs show the following results.

Of the five subreddits, /r/worldnews had the lowest amount of repairs confirming to reddiquette (19%) whereas /r/IAMA had the highest (55%). Of the repairs confirming to the reddiquette, the categories providing background and continuing the conversation were most prevalent. Respectively 54 (28.12%) and 42 (22%) instances of repairs could be attributed to these categories. They are clear outliers when compared to the rest of the categories. The third most prevalent category is thanking/condemning with 27 (14.06%) attributed repairs. Clarification and lexical errors follow with both 17 (8,85%) and 16 (8,33%) repairs. Retracting a statement had 7 repairs (3,64%), syntactical errors warranted 6 (3.12%) repairs, and repairs that indicated that an error was being corrected but not specified what type were made 6 (3.13%) times. Ambiguity reduction was found four times (2.08%) and live-blogging once (0.52%). Several categories in the table had no repairs corresponding to them, however there were 13 (7,29%) instances of repairs that couldn’t be attributed to any of the categories and were labelled as rest. Standing out here is the large difference in the totals of the overt repairs, ranging from 20 (/r/movies) to 55 (/r/worldnews) but all the subreddits tend to follow the general trend with repairs concerning the continuing of the conversation or providing background as the most prevalent.

**4.0.1 Examples of repairs.**

In the following tables we show some examples of the repairs. If a subreddit contained no repairs of a category it was omitted from the table.

<table>
<thead>
<tr>
<th>Type of repair</th>
<th>Subreddit</th>
<th>Example, preceded by “edit:”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguity reduction</td>
<td>Movies</td>
<td>which let me add I'm cool with Bruce Willis was just a jerk about it</td>
</tr>
<tr>
<td></td>
<td>ELI5</td>
<td>Terminator Edit: 2 (Note: this is regarding an excerpt from a Terminator movie. As there are several, the specific movie was added in the edit.)</td>
</tr>
</tbody>
</table>

Ambiguity reduction was found in two subreddits, in one instance the speakers message could be misinterpreted as sarcasm, and the post was edited to indicate that the speaker meant his post
sincerely. The second is more straightforward: there are several Terminator films, necessitating a repair in order for others to correctly identify the intended film.

In three subreddits repairs were found that are obvious error repairs, but cannot be classified without doubt as either lexical or syntactical errors. We've categorised these as uncategorisable errors, paradoxically as it may seem. A repair as “English is hard” may indicate a grammatical error, but is equally applicable to a misspelling. Other error repairs are more clearly identifiable.

<table>
<thead>
<tr>
<th>Type of repair</th>
<th>Subreddit</th>
<th>Example, preceded by &quot;edit: &quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error (uncategorisable)</td>
<td>Worldnews</td>
<td>Something something translation error.</td>
</tr>
<tr>
<td></td>
<td>News</td>
<td>a word.</td>
</tr>
<tr>
<td></td>
<td>ELI5</td>
<td>English is hard</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of repair</th>
<th>Subreddit</th>
<th>Example, preceded by &quot;edit: &quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical error</td>
<td>Worldnews</td>
<td>typo.</td>
</tr>
<tr>
<td></td>
<td>Iama</td>
<td>fixed spelling</td>
</tr>
<tr>
<td></td>
<td>News</td>
<td>spelling</td>
</tr>
<tr>
<td></td>
<td>ELI5</td>
<td>word.</td>
</tr>
</tbody>
</table>

Here we see, that with the exception of /r/movies all subreddits have repairs explaining the correction of lexical errors. Why /r/ movies does not have any lexical error repairs could have several reasons: redditors could simply not care, or the errors are not made or not made severe enough to warrant a repair in the eyes of the redditors.

<table>
<thead>
<tr>
<th>Type of repair</th>
<th>Subreddit</th>
<th>Example, preceded by &quot;edit: &quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntactic error</td>
<td>Worldnews</td>
<td>So to be syntactically correct: as a <em>matter of</em> fact, <em>I think</em>, strip search is really never acceptable</td>
</tr>
<tr>
<td></td>
<td>Iama</td>
<td>phrasing</td>
</tr>
<tr>
<td></td>
<td>News</td>
<td>Grammar</td>
</tr>
<tr>
<td></td>
<td>ELI5</td>
<td>grammar</td>
</tr>
</tbody>
</table>

A similar form of short repair is seen for repairs of syntactic errors. Generally they consist of short statements, with the exception of the repair in /r/worldnews, which is more elaborate.
Clarification edits aim to make a post more understandable for the reader. In the examples for /r/worldnews, /r/news and /r/eli5 this is explicitly stated: “to clarify”, “just to clarify”, “to add clarity” respectively are seen.

<table>
<thead>
<tr>
<th>Type of repair</th>
<th>Subreddit</th>
<th>Example, preceded by &quot;edit: &quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarification</td>
<td>Worldnews</td>
<td>to clarify, arabs living in israel have had voting rights since the creation of the state. I think you're confusing arab-israeli citizens living in israel and arabs living in palestine or occupied territories. <a href="http://en.wikipedia.org/wiki/List_of_Arab_members_of_the_Knesset">http://en.wikipedia.org/wiki/List_of_Arab_members_of_the_Knesset</a></td>
</tr>
<tr>
<td></td>
<td>Iama</td>
<td>Ninja edit: it's personal, I don't give a dang if you like their catchy music.</td>
</tr>
<tr>
<td></td>
<td>Movies</td>
<td>Overall just hated the Ryan Reynolds part of the film. Actually thought over all Green Lantern was okay.</td>
</tr>
<tr>
<td></td>
<td>News</td>
<td>Just to clarify, I don't want anything horrible to happen. I just believe that with the way our government is becoming, it's only a matter of time before it seriously impedes on our human rights to the point where we have to take action or just lose all of our rights</td>
</tr>
<tr>
<td></td>
<td>ELI5</td>
<td>OP was edited after this post to make ELI5 part more ELI5, and to add clarity.</td>
</tr>
</tbody>
</table>

Examples of thanking and condemning. In these example we see the edit function being used to convey thanks for reaching the front page of Reddit, and a thanks for a reply getting a Gold status (indicating a post of high quality). Note also the smiley face emoticon in /r/Iama. This would cause the script to fail if UTF-8 wasn't set as the character set.
These examples of repairs are providing background to an ongoing story. They do this by providing links to relevant articles or even the US constitution. Additionally, we see an edit providing background by covering a different subject in ELI5.

<table>
<thead>
<tr>
<th>Type of repair</th>
<th>Subreddit</th>
<th>Example, preceded by “edit: “</th>
</tr>
</thead>
<tbody>
<tr>
<td>Providing background</td>
<td>Worldnews</td>
<td>Which is not even to mention, as the poster below points out, that you’re <a href="http://mobile.reuters.com/article/idUSBREA3P08820140426?irpc=932">lying through your teeth</a>:</td>
</tr>
<tr>
<td></td>
<td>Iama</td>
<td>I should have said...Nercrafteria is our minecraft server which is, itself, funded through crowd-funding (though it’s server-side minecraft crowdfunding so probably not right for Patreon.)</td>
</tr>
<tr>
<td></td>
<td>Movies</td>
<td>So I went ahead and bought it because I couldn’t wait. From scanning the movie the whole thing appears to be in the same aspect ratio. I couldn’t find anything that was different. I used this image as a reference: <a href="http://i.imgur.com/4rMNuUZ.jpg">http://i.imgur.com/4rMNuUZ.jpg</a></td>
</tr>
<tr>
<td></td>
<td>News</td>
<td>apparently people here have severely forgotten their constitution. <a href="http://sparkcharts.sparknotes.com/history/usgovernment/section2.php">http://sparkcharts.sparknotes.com/history/usgovernment/section2.php</a></td>
</tr>
<tr>
<td></td>
<td>ELI5</td>
<td>This has nothing to do with how birds do it, but if you want to know how humans can do it... Record yourself saying something with b’s and p’s in it. Then reverse it, and you will hear a word. For instance we used to do sausage. [...]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of repair</th>
<th>Subreddit</th>
<th>Example, preceded by “edit: “</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuing the conversation</td>
<td>Worldnews</td>
<td>Edit: Now I feel like shit. :( Conversation went as follows: 1426580163.0 As a person who is half French: Fuck you, Quebec. I'm fucking tired of you and your ways. Bunch of retarded cunts. 1426580163.0 <del>Why are you being such an asshole, dude?</del> Edit: Now I feel like shit. :( 1426580163.0 Naw, man. I’m Metis. Being in Quebec, I was treated like dirt. Looked down upon like I was less than garbage. My grandfather almost kicked ass a few times (rest his soul) when I was out in the city with him. I’m allowed to say fuck that province because I’ve suffered their ignorance. Fuck them, fuck their ways and fuck Quebec in general. I hope they burn in a slow burning fire</td>
</tr>
<tr>
<td></td>
<td>Iama</td>
<td>One more, have you ever considered writing songs about more mainstream subjects (Like love and being the best in the rap game) to try and appeal to a wider audience with your music? I really like most of the songs of yours I’ve listened to (particularly Oh J.K. Rowling and Strange Charm), but I’m curious as to where you want to go from where you are with your music.</td>
</tr>
<tr>
<td></td>
<td>Movies</td>
<td>You should all look through this thread and read this guys posts. So much rhetoric there’s not enough room for content.</td>
</tr>
<tr>
<td></td>
<td>News</td>
<td>I’m aware that I was responding to sarcasm. I was agreeing with the poster’s sarcasm and stating my reason for agreeing.</td>
</tr>
<tr>
<td></td>
<td>ELI5</td>
<td>So women can do whatever they want and have a get out of jail free card. Edit: Well, why don’t you cry about it. 1426524124.0Woo sexism 1426524124.0Boo hoo. 1426524124.0Whadya gonna do? Cry?</td>
</tr>
</tbody>
</table>

In these examples the edit function is used to continue the conversation. In the example from
/r/ama we see that the edit function is used to ask another question in a post. Normally, one would suspect that a new post is used for another question, but here the edit function is used. In /r/worldnews the redditor making the edit seems to be making somewhat of an apology. Note that Metis are aboriginal people in Canada¹⁰. These examples do appear to break the flow of conversation at first glance, as they require a reader to scroll back to the post containing the edit after having found/read the post the edit refers to. However, as we shall cover below, this doesn’t necessarily is the case.

<table>
<thead>
<tr>
<th>Table 11: Examples of retracting a statement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of repair</strong></td>
</tr>
<tr>
<td>Retracting a statement</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

When a redditor notices that he or she has made a mistake in a post, they can retract their statement via a repair. These repairs were found in all subreddits with the exception of /r/movies. Retracting a statement in this way allows a redditor to keep his original claim and refute it in the same post, even when the error is pointed out in one of the replies to his post.

Rather peculiar to say the least is the fact that whilst prominent in the preliminary research there were no examples of live-blogging in the data. In the preliminary research we came across many examples of live-blogging, especially in /r/news and /r/worldnews, describing ongoing situations and providing update after update by a single redditor. Yet there were no examples to be found in the data. As the data is a snapshot of Reddit at a particular point in time, it may very well have been that there were nothing going on at that time that warranted ongoing updates.

¹⁰ https://en.wikipedia.org/wiki/M%C3%A9tis_people_(Canada)
These are examples of edits conforming to the reddiquette, but otherwise unclassifiable or not common enough to warrant a separate category. Spacing can be a cause for repair, to increase readability for example, but wasn't found often enough to warrant a separate category. Other edits, like the one in /r/news were incomprehensible, even when looked at in the original post and its context.

5 Conclusion

In this paragraph we will begin by answering our research questions (5.0.1) followed by a complete summary of the thesis (5.0.2). After this we will finish with a discussion section and provide suggestions for further research.

5.0.1 Answers to the research questions

To recapture, we set out to answer the following research question: Are repairs used on Reddit in a manner unique to Reddit or do they adhere to the conventions as seen in either CMC or classic papers on repairs? To provide the best possible answer to this main research question we split it up into three sub-questions. These are:

1. Do the repairs in conversation in Reddit adhere to the properties of repairs attributed to them by for example Shegloff (2014), Levinson (1983) and Levelt (1983)?
2. What are the conventions for repair in conversation and text-based CMC and does Reddit follow these conventions?

If Reddit doesn’t adhere to 1 or 2, how and why have the users of Reddit adapted their use of repair to the specific communication options of Reddit?

Prior to answering these questions we made the following hypotheses: it is probably unlikely that Reddit follows the categorisation of Levelt, or adheres to the distribution of the categories found by Levelt. The reason for that is twofold. First there is the difference between the face-to-face communication Levelt based his work on and the text-based, computer mediated communication

<table>
<thead>
<tr>
<th>Type of repair</th>
<th>Subreddit</th>
<th>Example, preceded by &quot;edit: &quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rest</td>
<td>Worldnews</td>
<td>spacing</td>
</tr>
<tr>
<td></td>
<td>Iama</td>
<td>south africa&gt;&amp;; (&gt; south america</td>
</tr>
<tr>
<td></td>
<td>Movies</td>
<td>Aliens. That's basically Aliens. So there's a template</td>
</tr>
<tr>
<td></td>
<td>News</td>
<td>Edit: /s</td>
</tr>
</tbody>
</table>
that defines Reddit. Second, Reddit allows what is essentially a self initiated self repair after the post is placed, and made part of the conversation. Whilst it is possible to make a repair after an utterance has been said or written, this repair becomes a new turn or sentence. Reddit allows for repairs to be made either as an addition to the post containing the reparandum, or as an outright replacement of the post or reparandum. Despite our assumption that the use of repairs would not adhere to the classification of Levelt, we did assume that Levelts classification would be a proper starting point for our research.

Assuming the repairs do not adhere to Levelt (1983) we hypothesised further that users of Reddit either follow the conventions for repair in text-based CMC or that these users adapted their use of repair to the constraints and affordances offered by Reddit.

By mining and analysing data from Reddit, we come to the following conclusions and answers.

1. *Do the repairs in conversation in Reddit adhere to the properties of repairs attributed to them by Shegloff (1992, 2014), Levinson (1983) and Levelt (1983)?*

Totalling the edits confirming to Levelt and the added categories, we see that 16.65 % of the overt edits fit in the categories laid out by Levelt, excluding the rest category. 83.31% fits in categories either specific to Reddit, or is part of the rest category (due to rounding, both figures do not add up to 100%). Having categorised all the repairs adhering to the reddiquette rule of specifying a reason for an edit, it appears that the categories of Levelt are hardly used as a reason for an edit by the redditors. Repairs are more often used to convey meanings or for reasons far more elaborate than mere error correction, as laid out in Sanders (2013) and Schegloff (2013). For example; spelling and syntactic errors are indicated as a reason for an edit but not as often as continuing the conversation, a far more elaborate reason to repair a post. It could be the case that most errors in spelling and syntax are corrected in private, before the post is visible on the website rather than publicly afterwards. The category continuing the conversation appears to be a bit on the broad side, with conversation enders, one-liners and shout-outs all grouped together. However, none of these were prevalent enough on their own to warrant being set in a separate category and all three are means to an end: the redditor wants have another turn in the conversation, and does so by repairing his post with a additional part. The difference between *ambiguity reduction* and *clarification* is also large. *Ambiguity reduction* was hardly found, but the more elaborate *clarification* was often a reason for repairs. The examples seen in Levelt, Schegloff, Levinson and Sanders are all repairs made within the same sentence, rarely consisting of more than a couple of words. Clarification on Reddit can consist of several sentences. Here redditors are breaking with the relative brevity encountered in
face-to-face repairs. As there is already a substantial cost to editing a post; a redditor needs to read, or re-read a thread, click buttons on the site and type a message, and typing can be a rather time consuming activity, the redditor might aim to have as little additional repairs as possible and provide as much extra information in his repair as possible.

With not even a fifth of the repairs fitting in his categories, we cannot say that Reddit adheres to Levelt. However, we must stress that we were only able to categorize a small portion of the total of edited posts found in the raw data. Other edited posts, where the author did not provide a reason for his edit may confirm to Levelt's findings. We will address the research of these covert edits below. The primary focus on the repair and the subsequent categorization of Levelt provided us with an excellent starting point for this research. In our humble opinion best way to categorize the overt edits on Reddit is by using an already existing categorization and expanding on it with categories found in preliminary research. That way the categories unique to Reddit can be added and analysed, whilst the results are still comparable against other research regarding repairs. When we look at the findings of Schegloff (1992, 2013) and Sanders (2013), Sanders findings that repairs can be used to convey more elaborate intentions than mere error repair hold without a doubt for Reddit, to the point that we could only categorize the repairs found by creating categories that covered what we saw during primarily observation. Schegloff findings in Repair after next turn (1992), where he covers the phenomenon of multiple repair spaces and all the possible locations a speaker or listener has to initiate a repair hold for Reddit, his findings amplified by Herring (1999) as she mentions disrupted turn adjacency amongst other difficulties encountered in text-based CMC. Although we haven't made an in depth study of a Reddit thread containing repairs and their subsequent context, it would appear that when redditors edit their post in order to repair as a reaction to a post made by a different redditor they take disrupted turn adjacency for granted and may take initiate a repair on turns that are nowhere near their original post. Schegloff's (2013) concept of operations which also expand on the limited definition of repairs is also found in Reddit, to the point that his categories and the categories we named appear to overlap.

We conclude that whilst Levelt's categorisation of repairs is present and provides us with an adequate starting point to base our research on, repairs on Reddit differ substantially from his findings. On the other hand, repairs on Reddit, and the reasons for using a repair confirm to the findings of Schegloff (1992, 2013) and Sanders (2013) in the way that they are used as a way to convey an intention more elaborate than the definition laid out in Levinson (1983).
2. What are the conventions for repair in conversation and text-based CMC and does Reddit follow these conventions?

As pointed out by Herring (1999) and Jacobs and Garcia (2013), text-based CMC in the form of chat is fast paced. This fast pace provides users of chat with difficulties regarding the placement of their repairs. Repairs can easily get misplaced in the ongoing conversation. This displacement is not only reserved for repairs, as displacement is common to CMC. Solutions to this problem include specific addressing and the use of repair-specific markers such as beginning a message with an asterisk. Using a repair marker in this way is a convention in chat. The use of markers like an asterisk in a new post on Reddit, as one would expect in chat, was found, but not in such a number that concluding that Reddit adheres to that convention stands. That is, the use of an asterisk as a repair marker by redditors. To indicate that a post has been edited, Reddit automatically adds an asterisk to an edited post. Additionally, Reddit asks redditors to provide a reason as to why they changed their posts. If a redditor heeds this request, he or she goes beyond merely indicating that a repair has taken place and letting other readers decide where and how to apply the repair to its intended post but makes it easier and more explicit for other readers of the thread. Rather than being used as a repair marker, we found that the asterisk was used by redditors to create bullet lists or indicate movement. With regard to repair markers, Reddit does not follow the convention for chat that repairs are placed in separate posts with a repair marker. Reddit does follow the convention for using a repair marker, as a feature automatically added by the site every time a user edits his post. Even with repairs placed in the same post they are aiming to repair (contrary to chat's convention of a new line) Reddit asks its users to provide a reason as to why the post is edited, and thus moves beyond the convention of merely indicating the repair with a marker.

Our research showed that a high percentage of posts are either edited or repaired without giving a reason. The high number of edits to posts without a reason could suggest that redditors monitor their posts after completion and find that some sort of either edit or repair is in order. However, we cannot be certain of that without further research. Although Reddit has some chat-like characteristics it also shares characteristics of a message board or forums. On a forum, somewhat similar in design to Reddit a user has more time to construct a larger message, compared to the barrage of messages filling the screen in a busy chat room. In fast-paced chat it is more comprehensible to write several smaller messages, rather than one large one, as that way issues with the insertion of messages from others and other causes of displacement pose less of a problem. One of the differences between repairs in chat and Reddit is the fact that in chat once the message is
posted, it has become permanent. Conversations in chat can be re-read in order to form a coherent conversation, but an earlier written line cannot be altered once posted. By allowing revision of placed posts Reddit breaks with that convention. A consequence of allowing a redditor to change his post is that once a post is replied to, he can replace it with a completely different message. If a redditor wanted to “troll”, he could for example ask who would agree with him on a trivial topic and later change his message to something of a sexist or racist nature. By indicating that a post is edited and by asking redditors to provide a reason, Reddit copes with alterations of this nature by making it explicit that a post has been edited. Another consequence is that of increasing the displacement. We have seen that repairs can be initiated several turns after the turn in need of a repair. In chat messages are placed in chronological order and re-reading by scrolling up and selecting only the messages relevant to your conversation allows a participant to recreate the conversation more or less as it would have taken place in face-to-face interaction. Reddit makes it far more difficult to make sense of the conversation. Even when posts are being displayed chronologically, the post containing a repair still has to be paired with the post that served as the other-initiator of the repair or contain a explicit mention of it being a self-repair. Whilst in the latter case, pairing repair and reparandum can easily be combined, in the former case it requires new scrolling to conversation to pinpoint the proper post.

Allowing changes to a post is not unique to Reddit: many, if not all online forums have that option. Asking that users provide a reason for their alterations is less often found. All in all, Reddit and its users seem to pick and choose what conventions to follow. As Reddit is neither a chat, nor an online forum but a rather unique form of text-based CMC, one can not expect Reddit to strictly adhere to the conventions for either form of CMC.

3. If Reddit doesn’t adhere to 1 or 2, how and why have the users of Reddit adapted their use of repair to the specific communication options of Reddit?

Not adhering to conventions for repairs and CMC begs the question as to why redditors and Reddit would break with them. What stood out first was that redditors do not just use repairs for mere error correction but for far more complex operations in the way Schegloff (2013) mentions. Secondly, despite the possibility of displacement problems redditors repaired their own post in response to posts far further down in the conversation thread. What stood out in this research was the way redditors use repair function to continue their conversation. Where in previous research (e.g. Schegloff, 2013) repairs are seen as an interruption of “the progressivity of the talk” (Schegloff, 2013: 43.), redditors use repairs to continue the conversation or clarify their intentions without
breaking the flow of the conversation. This difference makes the way Reddit handles repairs quite unique. Edits are not seen as an interruption needed to correct an error of some sort, but as a useful tool that allows contribution to the ongoing conversation.

Redditors also use the repair function to do far more than just repair their post. They combine the repair with the technological features at their fingertips. Rather than whatever knowledge is in their mind, they have the entire Internet at their disposal and they use it to provide sources for their claims, or to provide background information or context to a difficult subject.

Although prominent in the preliminary research, we were unable to find an example of live-blogging. The reason for this difference isn't obvious. Perhaps not all of the chosen subreddits or their topics lend themselves for live blogging but even in the subreddits that covered live or ongoing events during observation and preliminary research, live-blogging wasn't found in the research data. Despite the fact that we did not find examples of live-blogging in the research data, the fact that we found examples of live-blogging during observation prior to the gathering of the data and even during observation after having gathered the data to allows us at least the preliminary conclusion that redditors use the edit function in ways that drastically differ with the way repairs are used in conversation.

Repairs are even used in a social way, by editing a prominent posts to include a thank you to the other redditors who gave upvotes, or to thank the redditors for their participation. Examples of redditors using repairs to convey contempt for downvoters were also found.

Why would redditors adapt their use of the repair function in such a way? The answer may lie in the way the posts are displayed by default and the displacement problems that may occur because of that order, combined with the fact that posts below a certain threshold are not visible by default. Remember that replies to a post made in a thread are sorted by default on “best”, meaning that the best scoring posts are placed higher in the thread. It therefore makes sense for redditors to write their messages in such a way that they score the most points possible. It appears to be the case that when a redditor repairs his post via a reply, that this reply may not receive enough points to be visible. The post may also be so displaced by intermittent replies that it is buried between other replies, again not visible for other readers, despite it being a valid addition or correction. Redditors adapted to this limitation imposed by Reddit and used the edit function to not only repair errors (in the broadest sense of the word) but also as a way to make small extensions, or even add complete blog-like parts, to their posts. Their posts remains prominently displayed because of its already high enough score but now with an extra in the form of a repair. Using repairs in this way makes full use of the fact that a reader can back-trace the posts and re-read them to make sense of it. Like the tailor Schegloff (2013) mentioned, redditors seem to use repairs more to make alterations to their post to
improve it rather than to correct it.

The traditional view of repairs, as for example laid out by Levelt, indicates that repairs are there solely for the purpose of correcting an error: a slip of the tongue, a dis-ambiguous utterance or a variety of other reasons. There is some part of the utterance that needs repair, and a corrected, repaired new part of the utterance. These errors were seen as interruptions of the conversation, hindering its “flow”. Reddit is a form of CMC, based on written text and as such differs from conversation, obviously. Less obvious is the fact that the way repairs are used on Reddit differ greatly from the way repairs are used in conversation. Redditors use the fact that they are able to edit and repair their own posts in a unique way. Rather than seeing the edits as a nuisance, the redditors embraced the possibilities it brought beyond repairing an error. Limited by the default display of posts, redditors adapted their way of repair use to make statements that otherwise would probably go unnoticed. Not breaking the flow, but taking advantage of the strength of written text: the ability of a reader to read and re-read text. Instead of repairing an error redditors use the edit function of Reddit to create more elaborate repairs for more elaborate purposes: to create new sentences in a prominent post, to provide sources, to say thanks, give background information or even as a live-blog.

5.0.2 Summary

In this thesis we researched the way redditors use repairs in their messages. Reddit allows users to not only post messages, but also to edit them after completion. An edit can be made for a variety of reasons, repair being a prominent one. We were able to analyse the posts and provide an overview of the use of repairs on Reddit. To the best of our knowledge there is relatively little research on Reddit, with research regarding the use of repairs on Reddit non-existent. We have therefore attempted to provide an overview via a comparison with the findings on repair use by several of the prominent researchers on repairs. Starting with Levelt, as his focus lied predominantly on the repair itself, rather than on the conversation in which it occurred or the impact it had. By beginning at the repair, rather than an in depth analysis of complete conversation threads on Reddit itself we had a starting point to build on. By using the API provided by Reddit we were mine complete threads and extract all the edited posts. This data was then refined to contain only the posts confirming to the reddiquette requiring a redditor to provide a reason for his edit, so that we had a clear picture of the changes made in the post. Reddit only indicates that a post has been edited, and asks the redditor to provide a reason why, edits to posts aren’t logged so any categorisation along the lines of Levelt had to be done manually. We've categorised the reasons given by redditors for their repair according to
the categories Levelt used in his research and added our own categories based on our own observation of Reddit and on preliminary research of the literature concerning repairs in both conversation and CMC. We took our data from a selection of popular subreddits that characterised themselves with ongoing discussions rather than being a mere collection of one-liners. After refining we found that whilst Levelt proved to be a good starting point by large, redditors do not adhere to his distribution of reasons. Redditors were far more often found to use repairs in far more complex way than Levelt described. They use repairs not primarily for error correction, but as a way to share information or continue a conversation, apparently without interrupting the conversation in progress. In this regard they are breaking with the idea that repairs interrupt or break the ongoing conversation as is the case in spoken interaction. Redditors may use repairs in this way to correct possible displacement problems that may occur due to the insertion of other posts, or because of the way posts are ranked by other redditors and subsequently displayed or even hidden from view. By using repairs in this way, redditors are fully utilizing the affordances text-based CMC has to offer the ability to re-read as the reader sees fit and so be independent of a strict adherence to time and position of turns in a conversation. They do so in a manner unique to Reddit, breaking with several CMC-conventions. Although reddiquette suggests redditors also use a repair marker in the shape of “edit:” followed by their reason for a change to the post, this was only done in a minority of the edited posts.

5.1 Discussion and suggestions for further research

We’ve conducted our research on the use of repairs on Reddit. This research isn’t free from complications or debatable choices. In paragraph 5.1.1 we will critically examine the choice of the source-material and the choices made. Additionally we will provide suggestions for further research in 5.1.2

5.1.1 Critical remarks regarding the research

This research aimed to provide an overview of the use of repairs used on Reddit. When looking at the research the following needs to be kept in mind.

Reddit has a gargantuan amount of subreddits. It is possible, though unlikely, that the subreddits chosen, aren’t a proper representation of the way repairs are used on the entirety of
Reddit. Without broader mining and categorisation of the data we cannot be sure. Mining Reddit in its entirety (or a larger selection of subreddits) requires an even larger investment in time and a different approach in categorization of the repairs: rather than manual research, that process needs to be automated in some sort. The manual tagging of repairs was done by one researcher; the possibility of an error in tagging is possible. Even with automated tagging, the base-set could benefit from being created by more than one researcher.

The majority of repairs do not confirm to reddiquette as in many cases users do not give a reason for their repair. But as these repairs were not researched here, we must acknowledge the possibility that these repairs that do not resemble the results found in this research Researching these covert repairs would require a different type of research, perhaps with life-monitoring of a thread and recording the changes to a post before and after the edit. We must also acknowledge that the repairs researched here are not the first chance a redditor has to repair his posts: in composing his message, he can change wording, correct errors etcetera without being seen. Researching the way repairs are used in the composing of a message would require again a research that records the action of a redditor.

Finally, there is the script used to gather data, if it was able to indicate the different sub-threads in a post, that is; those that are indented on Reddit itself, it would be better possible to define the context in which the repair is made.

5.1.2 Suggestions for further research

*Hic sunt dracones*, Reddit is relatively uncharted territory when it comes to scientific research. Research in the past has focused on e-mail, chat, even role-playing games like Second Life and microblogs such as Twitter. Reddit has quietly sailed under the radar, even though it is an easily mine-able goldmine of conversation-data. The possibilities for further research are nearly endless and the following suggestions for further research no more than an attempt to shine some light on the tip of a very large iceberg.

**Finding the covert repairs.** By allowing repairs after the post has been made, Reddit provides its users with a second chance for repairs when compared to face-to-face communication. Many of the edits found did not adhere to the demands set for a usable, overt repair. These covert edits are a black box where erroneous text (at least in the eye of the redditor) goes in, is edited in some shape or form and placed back in the same place, without it being possible for an outsider to know what exactly was edited. As they make up approximately 70% of all edits found here, these covert edits make a difficult but interesting research target. How many repairs are made before a
post is placed? What shape do they have? By monitoring a number of redditors we might gain insight here. There are many practical caveats to this approach: how valid would the results be as participants have to enable monitoring of their online activity? On the other hand: is it ethically sound to monitor participants without explicit permission? Perhaps an experiment like this could be done in a classroom, with students making posts on Reddit and recording them. In a classroom, both obtaining permission and monitoring could be easier than recruiting participants and monitoring their own computers.

**An in depth case study is also an option.** We collected and analysed repairs on a large scale, and applied classification suited for this way of research. Reddits unique features could also be subject of an in depth case study: rather than gathering and categorising large amounts of data, a in depth case study could focus more on interaction in a specific subreddit or threat. The focus could then for example lay on the position of the repair in the conversation and whether or not the initiative for the repair came from the original poster, or was brought on by the incentive of others. The turns in other-initiated self repairs can be spread apart over a number of posts, instead of being close together as one would expect in face-to-face communication. What influence does this have?

**Repair use and a redditors accountability on Reddit is also worth researching.** In this thesis, the emphasis lies on the use of repairs, rather than its effect on a conversation. Rather than the effect a repair may have on the accountability of a redditor. Do repairs and edits influence the accountability or reputation of a redditor? Does frequent use of repair have an effect on the reliability of an individual redditor? More specific than the accountability of a specific redditor would be examining an individual post. Does repairing a post have an effect on the popularity of an individual post, and what about the popularity of that post the thread in which the post was repaired in a subreddit? A case study in a similar vein as the one done by Antaki et al. (2005) may provide us with more information on this issue.

**What about broken conversation pairs?** We have seen that there are many examples of repairs within a post, but far fewer repairs made as a reply to a post. In the example we saw in 2.5.1 regarding the winters in Toronto there was already one post by another user between the error and the repair. Displacement is an issue in CMC, even in Reddit where conversations are graphically managed to provide coherence. Howes, Healy and Mills (2009) have done research on split utterances in chat, and whilst at first glance not applicable to Reddit, their research method based on the monitoring of interaction could be adapted to monitor ongoing interactions in a browser set to Reddit. The monitoring of willing participants, being free to interact as they please (even with the limits they might impose on themselves due to being monitored in a classroom setting, or given an assignment to complete) can provide researchers with useful raw data on how a relatively
spontaneous conversation on Reddit takes shape. This research set-up has the added benefit that at least one side is being monitored during the creation of their messages and interaction with other redditors.

**Breaking the convention.** Reddit has its own jargon and conventions. Even with free-speech as an absolute given, subreddits can and do enforce stricter rules and moderators can ban users who break the rules. But what about breaking the conventions before moderators have to intervene? Is there some sort of social control, where (new?) redditors are being told by others that their behaviour is frowned upon? What shape does it take? Is it similar or different to other forms of “trolling” found on the Internet? Is Reddit a self-controlled social medium or does it have more of an “anything goes”-like attitude where (social) control is more lax? Singer et al. (2014) showed that Reddit is a more and more a self-referential community, but can this be reflected in the language used on the website?

Above examples hopefully clarify that Reddit is more than just another place on the internet for sharing cat-pictures or sharing headlines with the world. Whilst a popular site on the internet it appears to not so well known to the general public. Seeing how much information Reddit can provide, and how little of it is researched it is somewhat saddening that an enormous amount of information is being left untouched.
Literature


S. C. Herring, “Humor that Binds: Joking and Coherence in Internet Relay Chat”, Paper presented


Reddit f (2015) Index – IAmA https://www.reddit.com/r/IAmA/wiki/index#wiki_3._what_topics_are_and_are_not_allowed.3F [WWW]


