



D2.4 Qualitative Analysis of Rumours, Sources, and Diffusers across Media and Languages

Arkaitz Zubiaga, Peter Tolmie, Maria Liakata, Rob Procter
(University of Warwick)

Abstract.

FP7-ICT Collaborative Project ICT-2013-611233 PHEME
Deliverable D2.4 (WP2)

This deliverable describes the work conducted in Work Package 2 for the study of conversations around rumours in social media. It describes our efforts towards developing an annotation scheme for analysing each of the messages that participate in rumourous conversations, and we put it into practice by creating a large-scale corpus of annotated tweets in English and German, including links to other media. This report also summarises our qualitative analysis of conversations around different types of rumours, looking at the diffusion across different media and languages. A sample of the corpus produced in this process is released along with this deliverable, which will be extended with the release of the whole corpus in December 2015.

Keyword list: rumours, veracity, annotation scheme, social media, analysis

Project	PHEME No. 611233
Delivery Date	July 13, 2015
Contractual Date	June 30, 2015
Nature	Report
Reviewed By	Kalina Bontcheva, Thierry Declerck
Web links	http://www.pHEME.eu/
Dissemination	PU

PHEME Consortium

This document is part of the PHEME research project (No. 611233), partially funded by the FP7-ICT Programme.

University of Sheffield

Department of Computer Science
Regent Court, 211 Portobello St.
Sheffield S1 4DP
UK
Contact person: Kalina Bontcheva
E-mail: K.Bontcheva@dcs.shef.ac.uk

MODUL University Vienna GMBH

Am Kahlenberg 1
1190 Wien
Austria
Contact person: Arno Scharl
E-mail: scharl@modul.ac.at

ATOS Spain SA

Calle de Albarracin 25
28037 Madrid
Spain
Contact person: Tomás Pariente Lobo
E-mail: tomas.parientalobo@atos.net

iHub Ltd.

NGONG, Road Bishop Magua Building
4th floor
00200 Nairobi
Kenya
Contact person: Rob Baker
E-mail: robbaker@ushahidi.com

The University of Warwick

Kirby Corner Road
University House
CV4 8UW Coventry
United Kingdom
Contact person: Rob Procter
E-mail: Rob.Procter@warwick.ac.uk

Universitaet des Saarlandes

Campus
D-66041 Saarbrücken
Germany
Contact person: Thierry Declerck
E-mail: declerck@dfki.de

Ontotext AD

Polygraphia Office Center fl.4,
47A Tsarigradsko Shosse,
Sofia 1504, Bulgaria
Contact person: Georgi Georgiev
E-mail: georgiev@ontotext.com

King's College London

Strand
WC2R 2LS London
United Kingdom
Contact person: Robert Stewart
E-mail: robert.stewart@kcl.ac.uk

SwissInfo.ch

Giacomettistrasse 3
3000 Bern
Switzerland
Contact person: Peter Schibli
E-mail: Peter.Schibli@swissinfo.ch

History of Changes

Version	Date	Author	Changes
1.0	17.06.2015	Arkaitz Zubiaga	first version of the deliverable, some sections to flesh out
1.1	24.06.2015	Arkaitz Zubiaga	complete version of the deliverable, with the analysis to be expanded
1.2	03.07.2015	Peter Tolmie	background in conversation analysis revised, and qualitative analysis expanded
1.3	12.07.2015	Rob Procter	General edits
1.4	12.07.2015	Maria Liakata	General edits
1.5	13.07.2015	Arkaitz Zubiaga	Comments from reviewers addressed

Executive Summary

This deliverable D2.4 describes the work performed at the University of Warwick within the Work Package 2 (WP2) as a member of the PHEME project. Pursuing the goal of studying the spread of rumours in social media, as well as the way users discuss them online, the deliverable outlines the efforts to define an annotation scheme, create an annotated dataset of rumourous conversations in social media, as well as to perform a qualitative analysis. The annotation scheme described in this deliverable has been developed through an iterative process, with two rounds of validation tests with local experts and a new validation test through crowdsourcing. The validation tests have been carefully performed by looking at real data, including rumourous conversations collected from the microblogging service Twitter. The annotation scheme is intended to encompass the wide variety of types of rumours that may spread and be discussed in the context of different events and situations. These manual annotations enable us to perform analyses to gain insight on how rumours propagate in social media.

The development of the annotation scheme is informed by findings from the related disciplines of Conversation Analysis and Ethnomethodology, but also takes account of the specific characteristics that make social media different from face-to-face communications.

After coming up with the final version of the annotation scheme for social media rumours, it has been applied through crowdsourcing to a large sample of 330 conversational threads consisting of 4,842 tweets in English and German. This document describes the process through which we produced this dataset as well as the outcome. It also describes the qualitative research performed within Task 2.1, which sheds light on the discussions produced around rumours in social media. Along with this deliverable, a subset of the resulting dataset and annotations has been released, to be followed by a complete release of the dataset in month 24.

The dataset produced in this work package will also be utilised in other work packages by other members of the PHEME project, such as in WP3 and WP4.

Contents

1	Introduction	3
1.1	Relevance to PHEME	4
1.1.1	Relevance to project objectives	5
1.1.2	Relation to other work packages	5
1.2	Outline of the Deliverable	6
2	Defining and Characterising Rumours	7
2.1	Defining Rumours	7
2.2	Characterising Rumour Types	9
2.2.1	Rumour Types by Accuracy	9
2.2.2	Rumour Types by Acceptability	10
2.3	Actor Types	11
3	Conversation Analysis and Ethnomethodology	12
3.1	The Origins of Ethnomethodology & Conversation Analysis	12
3.2	Respecification	13
3.3	Respecifying Rumour	15
4	Related Annotation Schemes	16
4.1	Rumour Types	16
4.2	Factuality and Sources	17
4.3	Actor Types	18
4.4	Other Annotation Schemes for Conversations	19
5	A Dataset of Social Media Rumours	21
5.1	Building a Dataset of Rumours and Non-Rumours	21
5.2	Complementing the Dataset with Conversations, Information Flows, and Unpacking URLs	23
5.2.1	Complementing with Conversations	23
5.2.2	Complementing with Information Flows	24
5.2.3	Complementing with Content from External URLs	25
6	Defining a Crowdsourcing Methodology	26
6.1	Why the Annotation was Crowdsourced	26

6.2	Disaggregating Annotation Task into Microtasks	27
6.3	Crowdsourcing Tasks Parameters	28
7	An Annotation Scheme for Rumours	30
7.1	Preliminary Annotation Scheme	30
7.2	Validation and Revision of the Preliminary Annotation Scheme	31
7.3	Final Validation and Revision of the Revised Annotation Scheme	37
7.3.1	Dataset Sampling for Testing the Scheme	37
7.3.2	Validation through Crowdsourcing and Reference Annotations	38
7.3.3	Analysis of of the Crowdsourced Annotation	38
7.4	Final Annotation Scheme	41
8	Dataset Annotation	42
8.1	Dataset Sampling for Crowdsourced Annotation	42
8.2	Annotation of Rumourous Conversations	43
8.3	Outcome of the Crowdsourced Annotation	44
9	Extending Dataset with Rumour & Actor Types	47
9.1	Inferring Rumour Types and Actor Types	47
9.1.1	Determining Rumour Types	47
9.1.2	Categorising Users by Actor Type	48
10	Extending the Annotation Scheme	50
10.1	Moving towards annotation grounded in microblog analysis	50
10.1.1	The turn-taking mechanism	51
10.1.2	Topic	54
10.1.3	The organization of conversation as applied to tweets and the organization of tweets when seen as conversations	58
10.1.4	The intersubjective constitution of tweeting as a phenomenon	68
10.1.5	Following and followers	69
10.1.6	Tweeting as a mode of communication	69
10.1.7	Looking at microblogging as its own job of work with its own grammars of action	70
10.1.8	The asynchronous character of microblog exchange	72
10.2	The organisation of rumour as a feature of microblog exchange	74
10.2.1	'True' rumours	75
10.2.2	'False' rumours	78
10.2.3	'Unverified' Rumours	80
10.2.4	Speculation	82
10.2.5	Controversy	84
10.2.6	Agreement	85
10.3	Conclusion	86
11	Discussion	88

Chapter 1

Introduction

While inaccurate and questionable information has always been a reality, the emergence of the Internet and social media has increased this concern due to the ease with which such information can be spread to large communities of users [Koohang and Weiss, 2003]. This kind of information often starts as a rumour being posted by an individual on social media such as Twitter¹, Facebook², or Instagram³, and subsequently being passed on through their social networks and reaching a larger audience. The spread of rumours may have undesirable consequences as they can convey wrong information to people. Not only does this affect ordinary individuals who might pass on information without verifying it, but also professional practitioners such as journalists who may pick up a story from social media and inadvertently disseminate inaccurate or false information via news media. Given that the spread of inaccurate information can have dangerous consequences for society, the analysis of rumours becomes crucial to prevent the diffusion of inaccurate information and to identify information that is well backed up and verified.

The study of the spread of rumours in social media is attracting increasing interest within the scientific community [Friggeri et al., 2014, Hannak et al., 2014]. However, these studies have generally focused on virality and social network analysis of rumours and have not looked in more detail at the nature of rumours, how they are linguistically crafted, and how they are subsequently supported and/or denied by others in social media. We intend to fill this gap by first introducing an annotation scheme, a framework for systematic annotation of different aspects reflecting the content of rumours. Annotated datasets resulting from this scheme will assist to perform content-based studies on conversations around rumours, and to develop a system that automatically processes rumours texts in social media, as well as conversational aspects such as reactions around them.

One of the proposed ways of handling the development of an annotation scheme for Twitter feeds that moves beyond the work undertaken by [Procter et al., 2013b] on the

¹Twitter - <http://twitter.com/>

²Facebook - <http://www.facebook.com/>

³Instagram - <http://instagram.com/>

London riots, is to exploit existing work in the area of conversation analysis as a means of providing richer annotations of topics as they unfold. Here our primary concern will be to map out what a grounding of annotations in the microblogging domain might look like. In particular, we will argue that, whilst conversation analytic approaches will serve well as a source of inspiration, it is ultimately going to be necessary to re-specify the interest somewhat as “microblog analysis” in order to steer around the potential dangers of missing the lived character of how people reason about tweeting as an activity in its own right.

In this document, we present a review of previous research developing annotation schemes that are relevant to PHEME, study their applicability to our context, and introduce our own annotation scheme, which has been iteratively tested and revised, developed for the specific purposes of analysing conversation around rumours in social media and being able to capture the sequential and nested nature of the interaction. This annotation scheme has then been put into practice for the crowdsourced annotation of a large-scale dataset of rumours posted and discussed in social media in the context of 9 different events. Rumours and their associated conversations were collected from Twitter while the events were unfolding. We have performed the annotation of 330 threads, in English and German, associated with the 9 events, originally collected from Twitter but also enriched with content from other media such as news media and blogs. We analyse the outcome of this annotation process so as to shed light on the nature of conversations produced by different types of rumours in social media. This analysis is the result of the conversation analytical social science research conducted within Task 2.1 of Work Package 2 in PHEME. To perform this analysis, we qualitatively examine a small subset of annotated examples in order to understand how some parts of the approach we have been advocating may be brought to bear in order to further enrich our understanding of just what might be going on over the course of the production and spread of such tweets.

Along with this deliverable, we also release a subset of the annotated dataset, which will be expanded by releasing the whole dataset in month 24 (December 2015). This initial sample includes 10 threads, 8 in English and 2 in German, along with all the tweets annotated with the scheme, information flows, and media links. The dataset released in this deliverable is in turn an enriched dataset produced from a subset extracted from the rumour dataset released in WP8 [Wong-Sak-Hoi, 2015]. While WP8 produced a dataset of around 7,500 tweet threads annotated as rumours or non-rumours, here we release a smaller subset with 330 of those threads, further annotated for conversation analytical purposes.

1.1 Relevance to PHEME

This section describes the relevance of this deliverable to the PHEME project’s objectives, and how it relates to the other work packages in the project.

1.1.1 Relevance to project objectives

This document outlines our efforts in undertaking the objectives defined in the description of Work Package 2 (WP2). The goal of this work package is threefold: (1) development of an annotation scheme that enables the analysis of conversational aspects of rumours spread through social media, (2) building datasets of social media rumours by making use of the annotation scheme to annotate rumourous conversations, and (3) performing a qualitative social science analysis of the resulting datasets of rumours. The deliverables within this work package will be relevant to numerous other work packages, which will be using the datasets for different analyses, as well as relying on the analysis presented here for further understanding of rumour characteristics. When it comes to the use of the datasets, for instance, WP3 will be using it for the development of machine learning tools that link cross-media and cross-language rumours, and WP4 will be using it for detection of rumours and determining the veracity of rumours, among others.

The datasets created in this work package will also enable the study and development of a methodology and tools for the linguistic analysis of rumours using natural language processing. In order to carry out this research, here we: (a) perform an introductory study by characterising social media rumours, (b) perform an interdisciplinary analysis drawing in particular on Conversation Analysis and Ethnomethodology and (c) develop an annotation scheme which can be used for the automatic processing of rumours.

1.1.2 Relation to other work packages

The work presented here, especially the datasets produced by making use of the annotation scheme, will be used in various work packages for different objectives within PHEME's scope of studying rumours in social media. In Work Package 2, they will support the ontology modelling Task 2.2. Work Package 3 will deal with the development of open source methods to track the flow of rumours, where the corpora will be used for development, parameter tuning and initial evaluation. In Work Package 4, they will support LOD-based reasoning about rumours. Task 4.3 also discusses rumour types but their focus is on belief classification and information diffusion. Work Packages 7 and 8 will also deal with the annotation of corpora in Tasks 7.2 and 8.2. These tasks will annotate rumours of interest to the healthcare and journalism use cases, respectively, making use of the annotation scheme we have defined.

The work conducted in this work package has been developed largely in collaboration with SWI. The annotated datasets of rumours and non-rumours described in D8.2 [Wong-Sak-Hoi, 2015] is used here as an input for our study, which we further expand, annotate, and analyse. The development of the expanded datasets is crucial for the annotation scheme described in this deliverable which is in turn key for the subsequent creation of corpora of social media rumours, which will include annotations provided by human coders. The annotated datasets obtained through this process will then be used to conduct

research on social media rumours, as defined in the WP2 of the PHEME's Description of Work. The annotation scheme is also designed with the goal to facilitate subsequent computational analysis of rumours using machine learning.

1.2 Outline of the Deliverable

This deliverable is organised in the following chapters. Next, in Chapter 2 we provide a formal definition of rumours, which combines previous definitions from both scientific literature and dictionaries, and then we delve into the different rumour types as well as different actor types that participate in rumour diffusion. The following two sections provide some background relevant to our work. Chapter 3 outlines ideas from the fields of Conversation Analysis and Ethnomethodology and their relevance to the investigation of rumour. Chapter 4 discusses existing annotation schemes that are relevant to the purposes of PHEME. Before we get into the details of our main work, we first describe the rumour dataset which we rely on and complement for our purposes in Chapter 5. Then, we explain how the annotation work we have conducted has been crowdsourced and why in Chapter 6. Our proposed annotation scheme is then introduced in Chapter 7, describing how we have come up with a final version of the annotation scheme after an iterative round of testing and revising it. Then, after describing the annotation work we have conducted through crowdsourcing in Chapter 8, we explain the process we followed to finalise the datasets by including rumour and actor types in the annotation in Chapter 9. We discuss further the extension and iterative refinement of this annotation scheme, and present the qualitative analysis of social media rumours in Chapter 10. Finally, we summarise the conclusions drawn from this work, and discuss its limitations and our future work programme in Chapter 11.

Chapter 2

Defining and Characterising Rumours

A compelling annotation task requires first of all to come up with a solid definition and characterisation of rumours, so that we can properly inform the task. To this end, we first review existing definitions of rumour and put them together into a new one. Then, we present a typology of rumours and describe their characteristics. And to conclude the section, we describe the typology of authors that can participate in rumourous conversations.

2.1 Defining Rumours

While there is a substantial amount of research around rumours in a variety of fields ranging from psychological studies [Rosnow and Foster, 2005] to computational analyses [Qazvinian et al., 2011], defining and differentiating them from similar phenomena remains an active topic of discussion within the scientific community. Some researchers have attempted to provide a solid definition and characterisation of rumours so as to address the lack of common understanding around the specific categorisation of what is or is not a rumour. [DiFonzo and Bordia, 2007] emphasise the need to differentiate rumours from other similar phenomena such as gossip and urban legends. They define rumours as “*unverified and instrumentally relevant information statements in circulation that arise in contexts of ambiguity, danger or potential threat and that function to help people make sense and manage risk*”. This definition also ties in well with that given by the Oxford English Dictionary (OED): “*A currently circulating story or report of uncertain or doubtful truth*”¹. Further, [Guerin and Miyazaki, 2006] provide a detailed characterisation of rumours, emphasising what differentiates them from urban legends and gossip (see Table 2.1 for the characterisation of rumors, gossip, and urban legends). From the differences posited by these authors, we highlight the following:

- It is of general interest to most listeners.

¹<http://www.oxforddictionaries.com/definition/english/rumour>

	Rumours	Urban legends	Gossip	“Serious knowledge”
Of general interest to most listeners	✓	✓		✓
Of personal consequence & interest to listeners	✓		✓	
Deals with person known to speaker or listener			✓	
Truth difficult to verify	✓	✓	✓	✓
Must be credible despite ambiguities	✓		✓	✓
Can be ambiguous	✓	✓		✓
Short or long?	Short	Long	Short	Short
Uses a story plot		✓		
Attention gained with horror or scandal	✓	✓	✓	
New or novel	✓	✓	✓	✓
Can be humorous		✓	✓	
Unusual or unexpected		✓	✓	✓

Table 2.1: Characterisation of rumours, gossip and urban legends, by [Guerin and Miyazaki, 2006].

- It is of personal consequence and interest to listeners.
- The truth behind it is difficult to verify.
- It must be credible despite ambiguities.
- It can be ambiguous.
- It tends to be a short story as compared to e.g., urban legends.
- It gains attention with horror or scandal.
- It has to be new or novel.

In contrast, urban legends are stories that are usually not credible or of personal consequence to the listeners, but tend to be more engaging and attention grabbing. Urban legends also tend to be longer stories. The main characteristic that differentiates gossip from rumours is that the former deal with persons known to the speaker or the listener. Both urban legends and gossip can be humorous, but that is not a feature that commonly characterises rumours.

Despite attempts to categorise them as different phenomena, [Guerin and Miyazaki, 2006] posit that all three – rumours, gossip and urban legends – *are merely ways of keeping a listener’s attention, and are not independently*

definable in themselves except for their particular conglomerate of conversational properties.

Summing up, here we expand on the OED's definition with additional descriptions from rumour-related research, which is richer and we argue more appropriate for our purposes within PHEME. We formally define a rumour as a **circulating story of questionable veracity, which is apparently credible but hard to verify, and produces sufficient skepticism and/or anxiety so as to motivate finding out the actual truth.**

2.2 Characterising Rumour Types

Despite coming up with a generic definition for rumours, there are different ways in which rumours originate and are spread and discussed. This is why we believe that, in order to perform a thorough analysis of rumours in social media, we need to categorise rumours into different types. To define a typology of rumours, we look at the two main dimensions that differentiate types of rumours, i.e., the accuracy of the information presented in the rumour, and the acceptability expressed by the recipients. For each of these two dimensions, we list the different types of rumours we have defined, and delve into the features that characterise them.

While the original categorisation of rumours as suggested in the PHEME Description of Work includes 4 categories (i.e., misinformation, disinformation, speculation, and controversy), here we explain how and why this categorisation has evolved into a two-level categorisation, with a slight variation on the original categories.

2.2.1 Rumour Types by Accuracy

While the accuracy value of a rumour is usually unknown to most people when it emerges, the rumour can evolve into a corroborated status as time goes by and new evidence comes up. With the emergence of new evidence and corroborations or debunks, rumours can change to a resolved status by proving it true or false. Alternatively, rumours can also remain uncorroborated, and hence unverified potentially for a long term or even forever. Based on this categorisation of rumours by accuracy value, we define and characterise the following types of rumours:

- **Accurate Information:** the rumour is eventually proven true either by reputable sources or by strong supporting evidence.
- **Unverified Rumours:** despite the best efforts of journalists or other professionals, the veracity of some rumours remains unresolved, which cannot be categorised as true or false due to the lack of evidence.

- **Inaccurate Information (Misinformation or Disinformation):** the information presented in the rumour is false. This can occur in the form of misinformation, where the author makes an honest mistake by spreading wrong or misleading information, or in the form of disinformation, where the author deliberately spreads misleading information. One of our initial plans was to be able to differentiate between misinformation and disinformation among all the inaccurate information. However, since the distinction of misinformation and disinformation lies in the intent of the author, and given the difficulty of determining what the intent of the author might be in most cases, we found this distinction unaffordable even for the human annotator, and so we limited to the identification of inaccurate information irrespective of intent.

2.2.2 Rumour Types by Acceptability

Apart from the accuracy of a rumour, the other dimension that we deem characteristic of rumours is the acceptability expressed by the recipients. How a rumour is perceived, accepted, and subsequently spread by those who hear about it can determine the reach of the story. In the specific case of social media, the acceptability of rumours can be observed in the replies and the subsequent tweets in the life cycle of a rumour. Building on the original categorisation set forth within the PHEME project, we define the following three types of rumours by acceptability:

- **Speculation:** rumours that can be categorised as speculation include early reports that lack supporting evidence. The OED defines speculation as “the act of talking a matter over conjecturally.”
- **Controversy:** controversial rumours are those that produce skepticism by having many recipients question the veracity of the story, by disagreeing or presenting opposing views. The OED defines controversy as “The action of disputing or contending one with another; dispute, debate, contention.”
- **Agreement:** cases where users do not question the veracity of a rumour, and all or most assume it is truthful, produce a high level of acceptability, which we refer to as “agreement”.

It is worth noting that the categorisation of rumours by acceptability does not depend on the accuracy of the rumour, and any value of acceptability can apply irrespective of the accuracy value of the rumour. Note also that while controversy and agreement are exclusive, speculation can apply together with one of these two to the same rumour, as a speculative rumour can in turn spark controversy or agreement. Later in the analysis we further elaborate on the approach we follow to categorise rumours as being speculative, controversial, or agreeing.

2.3 Actor Types

For the categorisation of social media users into different actor types (i.e. distinct categories of social media users), we look at different factors. These factors allow us to analyse what type of role they might play in a rumourous story. The three factors we look at include:

- **Verified vs non-verified users:** Twitter grants a special “verified” status to users whose authenticity has been checked. By having an account as “verified”, Twitter confirms the authenticity of the user’s identity, and it is mostly used for key individuals such as celebrities and well-known professionals, as well as for brands². The fact that a user has been verified or not therefore provides a reputation level that we want to analyse in the context of rumours.
- **Followers / Following ratio:** the number of followers a user has is a value that shows the importance and reputation of the user. The more followers a user has, the more likely their credibility and reputability. However, it is also important to consider how many accounts a user is following (number of followees), since some users simply follow others to get more followers and boost their reputation. Hence, by defining the followee to follower ratio, we express the difference in terms of orders of magnitude between the number of accounts a user is following and the number of accounts following the user.
- **Journalist or news organisation:** different from regular users, journalists and news organisations usually have the commitment to make sure that the information they post is accurate and has been corroborated. However, they also make mistakes occasionally. To study how journalists and news organisations participate in conversations around rumours, we differentiate them from other users so we can analyse their behaviour.

These factors are independent of one another, and hence every user will have a value for each of the factors. We will use these to analyse how they might affect the role of a user in the conversational aspects of determining the veracity of a rumour. In the analysis part of this deliverable, in Chapter 9, we further detail how we categorise authors.

²<https://support.twitter.com/articles/119135-faqs-about-verified-accounts>

Chapter 3

Introducing the Conversation Analytic and Ethnomethodological Approaches

The preceding section of this deliverable is concerned with arriving at a workable definition of 'rumour' for the purposes of informing the collection of tweets that are immediately identifiable as rumours for the purposes of annotation. However, a longer-term strategy that we shall also be adopting in Work Package 2 is the use of Conversation Analytic and Ethnomethodological approaches to understand how tweets and whole bodies of related tweets are organised as social accomplishments, and how rumours are therefore constituted within this as social accomplishments in some way.

3.1 The Origins of Ethnomethodology & Conversation Analysis

Ethnomethodology first arose as an approach for analyzing social phenomena in sociology in the 1950s. Its principal articulation as a programme of research can be found in the works of Harold Garfinkel, most notably his *Studies in Ethnomethodology* [Garfinkel, 1967]. What might be termed a radical empirical sociology, it is heavily influenced by the phenomenological writings of Edmund Husserl and the later philosophical writings of Ludwig Wittgenstein. Its fundamental concern is with how orderly social phenomena are organised by people themselves through concerted local action to produce them as being recognizably the social phenomena they are taken to be. It focuses on what people do, and how people do it as a matter of method, such that everyone else can see that that is indeed what they are doing.

Conversation Analysis was first developed through the work of Harvey Sacks (see the *Lectures on Conversation* [Sacks, 1995]). Sacks was one of Garfinkel's students and his studies of conversation can be seen as a practical working through of Ethnomethodology's programme by taking a readily available phenomenon within society — namely 'talk' —

and seeing just what its organised properties might be as features of the social order.

In a seminal paper dating from 1963, Sacks made some important observations regarding the ways in which sociological description tended to be pursued at that time (and, for the larger part, since as well). To illuminate his concerns Sacks conceived of a machine at an industrial exhibition consisting of two parts where one part is designed to undertake some particular job whilst the other part systematically and contiguously provides a narration of what the first part is doing. He suggested that a lay understanding of this machine would be something along the lines of a ‘commentator machine’ and that any attempt to make sense of the machine would involve being able to reconcile the relationship between the parts doing the job itself and the parts doing the narration.

Sacks’s idea here is to use the machine to represent the social world where you have a whole bunch of stuff going on that together constitutes the ‘doing’ part of society but, at the same time, you also have a bunch of talk going on whereby people systematically narrate their lives, the ways in which their lives are organised, and through which many of the ‘doing’ parts get implicated or even done. For Sacks the point is that, to understand the social world you cannot split those two bits apart and make use of the narration part without first of all looking at the narration part to see just how that works as well. What he is alluding to here is the tendency within social science to make use of language imported from the commonsense, everyday world without first of all opening up to inspection the work that language does in the world. Social scientists make use of language unreflectively as a resource for doing the job of description of the social world without taking that use of language to be itself a topic for investigation. Thus social scientists are, in reality, just engaging in the same work as the other narration components within the machine.

Sacks’s work, and by necessity the rest of Conversation Analysis, is therefore heavily invested in the business of taking the social production of language-based phenomena as a serious topic for investigation in its own right. As tweets are also language-based phenomena with their own organizational properties, we are therefore similarly seeking to understand tweets in this kind of way.

3.2 Respecification

Alongside of this interest of Sacks in the problematic character of sociological description, Garfinkel had already been developing what he called a foundational ‘respecification’ of the problem of sociology. It was founded upon a re-working of Durkheim’s famous aphorism that “The objective reality of social facts is Sociology’s fundamental principle” [Durkheim et al., 1938]. Grounded in his own reading of the works of Edmund Husserl, Garfinkel sought to reframe this as a matter of it not being the case that there were just social facts out there to be picked up and inspected, so to speak, but rather that the sense of something counting as a social fact was something that was accomplished itself by the ordinary members of society. The problem was therefore opening up for

inspection what this accomplishment might consist in. People take for granted that there is order in the world and they expose in everything they do just what kinds of orderly arrangements they are presuming will hold. The job of the sociologist is to uncover and bring into view these assumptions about 'the way the world works' and to explicate the ways in which they provide methodologically for the production of orderly phenomena.

The notion of respecification was absolutely central to the work of Garfinkel. It can be seen to resonate through much of his writing but he gives explicit voice to the idea in several places. In an edited transcript of a conversation between Garfinkel and Benetta Jules-Rosette recorded in the summer of 1985 he presents respecification in the following way:

“Our studies developed a radical, alternate technology of social analysis. Some of its policies are well known ... These and others were developed in the attempt to avoid the intractable absurdities that everywhere accompany classic methods of analytic social studies of practical action. With our alternate methods we have specified several identifying issues of the problem of social order as discoverable phenomena in and as immortal ordinary society ... These identifying issues are only discoverable. They cannot be imagined and they cannot be obtained by operating on representations of social order. Their import is that they respecify the ordinary society and do so in inspectable, detailed ties between practical action and the phenomena of order/production.”, Garfinkel and Jules-Rosette, 1986, unpublished transcript

Using the placeholder 'order*' for all possible topics of interest 'in-and-as-of-the-workings-of-ordinary-society' Garfinkel offers a further articulation of the idea in another later volume of collected works [Garfinkel, 1991]

“Not only the topic of detail, but every topic order is to be discovered and is discoverable, and is to be respecified and is respecifiable, as only locally and reflexively produced, naturally accountable phenomena of order*. These phenomena of order* are immortal, ordinary society's commonplace, vulgar, familiar, unavoidable, irremediable and uninteresting 'work of the streets'.”*, [Garfinkel, 1991]

This notion of respecifying what it is we might be talking about and not taking for granted articulations of the social world as features of the social world but rather inspecting how people themselves make them a feature in some way, is central to our own longer-term strategy within PHEME.

3.3 Respecifying Rumour

So, whilst there is a pragmatic necessity involved in pinning down what phenomena count as 'rumour' for the realization of a workable annotation scheme, this should be set against a longer-term concern with not just taking these assignments for granted but rather treating them as ongoingly revisable according to what our investigations into rumour production in Twitter reveal about how people themselves reason about rumour in various ways. Thus we consider ourselves to also be involved in the job of taking what has been construed as a topic in other fields, namely 'rumour', and considering what it could amount to as a topic of investigation 'in-and-as-of-the-workings-of-ordinary-society'.

A feature of our work over time will therefore be a respecification of 'rumour' as a topic of interest for sociological investigation by focusing upon what the 'local production', 'natural accountability', and 'coherence' of phenomena conventionally glossed as [rumours] look like in praxis. To do this involves moving away from taken-for-granted assumptions about what 'rumours' might be, and towards what it might take to be able to call something a 'rumour' — reasonably or otherwise — in lived social action. It also involves exploring what work members of society are engaged in when they articulate the proposition that something might be a rumour. What kinds of things does ascribing something the status of a 'rumour' accomplish in the world? As a programme of work this will involve setting aside taken-for-granted notions of what rumour might amount to and instead looking at: a) what kinds of features of interaction are taken by members themselves to be recognizable as 'rumour' in some way; and b) what kinds of work in interaction ascriptions of rumour to phenomena might be seen to do. In particular this will be directed towards an examination of how rumour-relevant phenomena are organised features of microblogging practices, and specifically the use of Twitter, in their own right.

This exercise will build upon an existing corpus of work in the conversation analytic and ethnomethodological literatures that already touches upon rumour-related matters in various ways. Relevant texts here include: [Meehan, 1989], [Mellinger, 1992], [Rapley, 1998], [Smith, 1978] and [Wooffitt, 1992] with regard to the accomplishment of 'facticity'; [Coulter, 1979], [Harper, 1994], [Jalbert, 1989], [Sacks, 1995] and [Sidnell, 2011] regarding 'belief'; [Antaki, 2000], [Bergmann, 1993], [Goodwin, 1980], [Parker and O'Reilly, 2012] and [Sacks, 1995] with regard to 'gossip'; and [Clifton, 2009], [Heritage et al., 2001] and [Sacks, 1995] regarding 'subversion'.

Chapter 4

Related Annotation Schemes

Here we discuss the most relevant annotation schemes and corpora that are closely related to the purposes of our work on the development of an annotation scheme for rumours. We have organised the annotation schemes into the following subsections: (i) rumour types, (ii) factuality and sources, and (iii) author types.

4.1 Rumour Types

[Procter et al., 2013b] conducted a study of tweets sent during the 2011 England riots. They grouped tweets into “information flows”, which is defined as a thread of tweets that retweet and make comments on a common source tweet. They looked at popular (i.e. large) information flows, and categorised them into an introduced typology of messages – media reports, pictures, rumours and reactions – as well as of author types. The paper provides detailed lists for both types of messages and authors. In the specific cases of rumours, they include the following subtypes: (i) claim without evidence, (ii) claim with evidence, (iii) counterclaim without evidence, (iv) counterclaim with evidence, (v) appeal for more information, and (vi) comment. They identified and characterised how rumours begin with someone tweeting an alleged incident, and quickly pick up popularity as others retweet and spread them. The veracity of a rumour is eventually questioned as Twitter users subject it to various “facticity tests” (e.g. questioning evidence, applying “common sense reasoning”) and over time a consensus is usually reached. However, the authors posit that even previously refuted rumours can re-surface and continue to be spread.

[Qazvinian et al., 2011] studied the automatic detection of rumours from tweets. They dealt both with retrieval of rumour-related tweets, as well as with identification of whether the tweet author endorsed the rumour. In the first step, they categorised a tweet as a rumour or non-rumour, whereas in the second step they categorised those deemed rumours as the author of the tweet confirming it, or denying/doubting/questioning the veracity of the rumour. They used some manually defined queries to retrieve tweets that potentially

concerned rumours (e.g., “Obama & (muslim—islam)” for the rumour on whether Barack Obama is muslim). They developed a classifier using three different types of features: content-based, network-based, and Twitter-specific features. They found that content-based features led to the best classification performance both for the rumour vs non-rumour and for the rumour support vs denial/questioning classification.

[Soni et al., 2014] investigated how linguistic resources and extra-linguistic factors affect perceptions of the certainty of quoted information on Twitter. They collected tweets posted by 103 American journalists and bloggers, which were identified from lists of journalists on Muckrack.com¹ and selected quoted content from those journalists by filtering tweets with source-introducing predicates (e.g., claim, say, insist) listed by [Saurí and Pustejovsky, 2012]. Then they used Amazon Mechanical Turk to annotate a subset of 1,265 tweets with quoted content from the journalists. The turkers rated the tweets in a 5-point likert scale from “Certainly False” to “Certainly True”. Using regression techniques, they studied the correlation of features with a claim being true or false. The features they analysed include: (i) cue words, (ii) cue word groups, (iii) source quoting the content, (iv) journalists as the authors of the tweet, and (v) claims as the bag-of-words of the text in the tweet. They found that cue words used to introduce the claim did correlate with the factuality perceptions, but other extra-linguistic factors such as the source and the author were not relevant.

[Zubiaga and Ji, 2014] relied on four aspects that determine how people perceive the veracity of a piece of information: (i) authority, (ii) plausibility and support, (iii) corroboration, and (iv) presentation. They conducted a study where users rated each of these four features for tweets and found that users mostly rely on author details to determine the veracity of a tweet, even though some author details such as location and description are not readily available on Twitter and third party clients’ feeds. Additionally, they found that corroboration often misleads viewers into falling for a hoax, misunderstanding that the existence of many supporting claims does not necessarily mean a rumour is true, which matches up with previous findings in Psychology research for offline information verification.

These existing annotation schemes for rumours have their merits, but are not detailed enough for our purposes. We will consider them in the development of our scheme, incorporating new factors in order to drill down further into the nature and salient features of rumours.

4.2 Factuality and Sources

[Saurí and Pustejovsky, 2009] described the annotation scheme as well as the process they followed to annotate the existing TimeBank corpus [Pustejovsky et al., 2003] with event factuality details. While TimeBank includes temporal and event information, FactBank

¹<http://muckrack.com/>

adds a new layer providing information about the factuality of those events. Event factuality considers two dimensions, polarity – positive (+), negative (-), or underspecified (u) – and modality – certain (CT), probable (PR), possible (PS), and underspecified (U). The annotation was performed by two students, who were instructed to ignore any kind of real world knowledge and annotate the content of the sentences. This annotation led to an inter-annotator agreement (computed on 40% of the corpus) of $\kappa_{COHEN} = 0.81$. They found the annotation to be skewed towards cases that were certainly positive (CT+) and underspecified (Uu), which was not surprising as the corpus was made of news articles and these types of statements would be expected to predominate. In addition, the annotation scheme also includes the *events*, which are part of the original TimeBank corpus, the *sources* mentioned in the statements and *other sources that are relevant* to the statement, such as the text author.

[de Marneffe et al., 2011] collected annotations through Mechanical Turk for the FactBank corpus, which in this case referred to the veridicality of the sentences, defined as the perceived likelihood of a piece of information being true, informed by context and real world knowledge. The turkers achieved a lower inter-rater agreement ($\kappa = 0.53$) than [Saurí and Pustejovsky, 2009] did with two annotators. They then built a maximum entropy classifier to automatically determine the veridicality of the sentences.

[Vlachos and Riedel, 2014] described the creation of a corpus of fact checked statements. Using statements PolitiFacts’ Truth-O-Meter² and the fact checking blog of Channel 4³ as sources, they curated a set of statements annotated as True, MostlyTrue, HalfTrue, MostlyFalse, and False (the two sources employ different categorisations of truth, which were manually combined). They removed all statements that could not be corroborated with online sources. The corpus includes 106 statements at present, which will be made available online⁴.

While the above annotations have been collected for news and political statements, which we could expect to be grammatically richer and more precise in terms of the factuality expressed, the annotation scheme could also be readily applicable to social media posts like tweets. It is likely that social media posts being grammatically less comprehensive would lead to more “underspecified” statements, which we will study in detail during the annotation process. Similarly, we expect that the source of a rumour in a tweet might not be as clear as in other texts such as news.

4.3 Actor Types

As in other forms of communication, the identity of the person posting (“authoring”) content on social media may have a bearing on how recipients assess its likely credibility.

²<http://www.politifact.com/truth-o-meter/statements/>

³<http://blogs.channel4.com/factcheck/>

⁴<https://sites.google.com/site/andreavlachos/resources>

For example, where there is knowledge of the poster's previous trustworthiness, this will influence how new postings are assessed. Similarly, where the poster is understood to be acting in a professional capacity (e.g., as a journalist), then this (and the organisation they represent) may also influence how postings are assessed.

[De Choudhury et al., 2012] researched the development of an automatic classification system that identifies types of users on Twitter, which can be useful to differentiate them in the context of events. They introduced a categorisation of three types of users, which included organisations, journalists/media bloggers and ordinary individuals. They used vectors represented by the following features for the classification: number of followers and followees, number of tweets posted, the fraction of tweets that are replies, the presence/absence of named entities and the topical association of the user's history from a list of 18 topics. The named entities and topics were derived using OpenCalais⁵. They use a kNN classifier, which empirically performed better than 9 other classifiers that they tried. Experimenting with tweets associated with 8 different events, their classifier performed most accurately when categorising ordinary individuals, with slightly lower performance values for journalists and organisations.

In their study on the spread of rumours in the context of the 2011 England riots, [Procter et al., 2013b] also introduced a typology of types of authors that posted the tweets. This typology included up to 20 types of authors, which defined a fine-grained categorisation, differentiating, for instance, ordinary individuals from rioters or from researchers. While this represents an exhaustive categorisation of users, it appears to be specifically crafted for riots and it might need to be revised to generalise it to other types of events.

Both of these annotation schemes for author types are of interest for our purposes when annotating authors in rumours. However, while the first might not be specific enough to consider all the author types that we might need to differentiate in the context of rumours, the second might need to group some of the types into higher level types to make it generalisable to a wider variety of event types.

4.4 Other Annotation Schemes for Conversations

There are other annotation schemes that also analyse conversational aspects of textual communication, but significantly differ from the purposes of PHEME of annotating rumours. For instance, some have made attempts to categorise types of dialogue that occur during argumentation. One such example is the categorisation made by [Walton, 2010], which includes seven types of dialogues that were observed in cases of argumentation: (i) persuasion, (ii) inquiry, (iii) discovery, (iv) negotiation, (v) information-seeking, (vi) deliberation and (vii) eristic. While this categorisation also deals with conversational practices, it clearly differs from rumours. Even though some types such as information-

⁵opencalais.com

seeking can also apply to rumours (here we define it as “*appeal for more information*” to code the way a statement is presented), other types like negotiation are not straightforwardly applicable to rumours. Related to this, in our own annotation scheme, described below, we initially included a feature called *presentation*, which was intended to code for the type of dialogue.

[Ritter et al., 2010] looked at the use of topic modelling approaches for categorisation of tweets within conversations. They identify conversations from Twitter as sets of tweets responding to each other. They list 8 types of conversational messages for Twitter: status, question to followers, reference broadcast, question, reaction, comment, answer, and response. While this is an interesting typology of conversational messages observed in Twitter dialogues, it is rather generic and does not specifically tie in within the context of social media rumours. For our annotation scheme, we define a similar typology for the specific case of rumours discussed in social media.

Chapter 5

A Dataset of Social Media Rumours

Before getting into detail describing the annotation scheme we have developed for social media rumours, and performing the annotations using the scheme, first we describe the data collection process we carried out. This step has been performed in close collaboration with SWI in their objectives within PHEME’s Work Package 8, and further details can be found on the deliverable D8.2 [Wong-Sak-Hoi, 2015]. Here we briefly describe the data collection process, and summarise the outcome, which is crucial for our subsequent steps.

5.1 Building a Dataset of Rumours and Non-Rumours

For the analysis of conversational aspects around rumours in social media, here we make use of the dataset built in collaboration with SWI in PHEME’s Work Package 8 [Wong-Sak-Hoi, 2015]. This annotation has been conducted by SWI, using the annotation tool that UWAR developed to that end (see Figure 5.1). The annotation tasks performed by SWI have enabled the identification of rumours and non-rumours associated with 9 different events, as well as additional annotation, which we briefly summarise next due to its relevance to our analysis. Through our annotation tool, we sampled a set of tweets for each of the 9 events; these tweets were sampled by selecting the most retweeted tweets, which allows us to select the tweets that sparked most interest, in line with our definition of rumours. The sampled source tweets, which needed to be annotated as rumours or non-rumours, were enhanced with additional context from the conversations associated with them. We describe how these conversations were collected in Section 5.2. Table 5.1 summarises the statistics of the annotation work, showing the number of rumours and non-rumours identified for each of the 9 events.

For the purposes of our study here, we focus on the 2,695 tweets annotated as rumourous. These rumourous tweets have been annotated in three different languages: 2,460 in English, 198 in German, and 37 in French. This allows us to perform a cross-lingual study. For these rumourous tweets, the annotation has also included additional

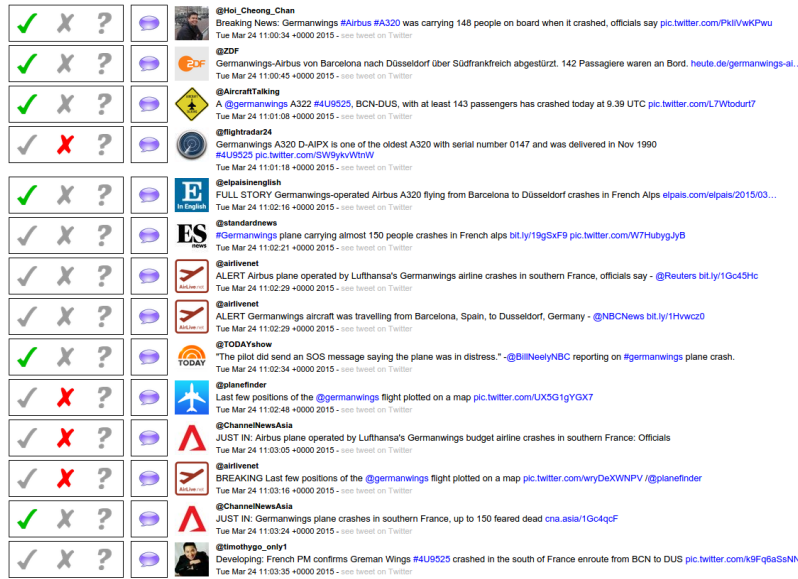


Figure 5.1: Screenshot of the annotation tool used by SWI to identify rumours and non-rumours.

annotations that are crucial for our study within Work Package 2:

- Categorisation of tweets within stories:** besides identifying whether the tweets could be deemed rumourous at the moment of posting, the manual annotation also categorised the rumours into stories. The categorisation of rumours into stories was performed by associating each story with a “rumour title”, which in turn allows us to have multiple tweets grouped within the same rumour title. With multiple tweets within a rumour title, we can build stories comprised of a timeline of tweets referring to the same rumour.
- Addition of media links:** each of the stories was also manually associated with

Event name	All threads	Annotated	Rumours	Rumour stories	Non-rumours
Sydney Siege	1966	1321	535	61	786
Ottawa Shooting	1152	901	475	51	426
Charlie Hebdo	2129	2169	474	61	1695
Germanwings	5282	1024	332	19	690
Ferguson	12595	1183	291	42	892
Prince to play in Toronto	298	241	237	6	4
Gurlitt	684	386	190	3	196
Putin missing	923	266	143	6	123
Essien has Ebola	18	18	18	1	0
TOTAL	25047	7509	2695	250	4812

Table 5.1: Outcome of the annotation of rumours (see D8.2 for further details).

additional sources on the Web covering the story. The association of web links with a rumourous story included the annotation of two more values: (1) whether it was an article from news media, a blog post, a social media post, or another type of media, and (2) the position of the article as supporting (for), observing, or denying (against) the veracity of the rumour.

- **Identification of rumours proven true, false (mis/disinformation), or remaining unverified:** each of the stories was further investigated by journalists at SWI, looking on the Web, multiple news media, as well as social media, to find out whether a rumourous story had been later proven true or false. In cases where the truthfulness of a story was later proven true or debunked, it was annotated as such. Otherwise, the story was not be marked for any of these two options, which reflects that the story remained unverified, to the best of our knowledge.
- **Annotation of turnarounds:** the many tweets that conform a story express different stances. With turnarounds, we wanted to capture the first tweet in the timeline of the story that introduces the first clear stance that either verifies the story as true or discredits it by proving its falseness. This is only applicable, by definition, to stories annotated as proven true or false. The manual annotation included the selection of the very first turnaround tweet that provides the first opposing view from a reputable source or supported with strong evidence.

For further details on this annotation process, please refer to D8.2 [Wong-Sak-Hoi, 2015].

5.2 Complementing the Dataset with Conversations, Information Flows, and Unpacking URLs

While the 2,695 rumourous tweets from the dataset described above and created within the Work Package 8 are used here as a starting point, for our purposes of performing different analyses around rumours, we need to complement the tweets with conversations, information flows, and external URLs, which we describe next. The resulting dataset including conversations, information flows and external URLs is used in subsequent Chapters of this deliverable.

5.2.1 Complementing with Conversations

We rely on the above dataset which provides what we call the rumourous source tweets, for which we collect the conversations they sparked. As a native feature on Twitter, users can reply to one another. Hence, we look for all the replies that came after the 2,695 rumourous source tweets for the 9 events in the dataset. While Twitter does not provide

an API endpoint to retrieve conversations sparked by tweets, it is possible to collect them by scraping tweets through the web interface. We developed a tool that enabled us to collect and store complete conversations for all the rumourous source tweets. As a result, the conversation sparked by a source tweet can be visualised in a thread, as shown in Figure 5.2.



Figure 5.2: Example of a conversation generated by a rumourous tweet.

The collection of conversations for the 2,695 rumourous source tweets obtained 34,849 tweets, 2,695 being source tweets and 32,154 being replies to those.

5.2.2 Complementing with Information Flows

While the source tweets are posted by a single author, and can be responded to by others with different types of thoughts and opinions, another important type of activity that we want to capture is the spread of the tweet, which we refer to as information flows. The term information flow within the context of Twitter was first coined by [Lotan et al., 2011], who define it as “an ordered set of near-duplicate tweets”. In practice, this can be achieved by putting together retweets of the source tweet, i.e., tweets with the original content of the source tweet, which are passed on by a user to their followers. Hence, for all 2,695 source tweets, we put together all the retweets available in our datasets. This amounts to 62,163 retweets for all those source tweets, an average of 23.07 retweets per source tweet.

5.2.3 Complementing with Content from External URLs

To enable the cross-media analysis that we planned, we also collect the content from other media, so that we complement our dataset originally focussed only on Twitter. The cross-media links were complemented in two different ways: (1) by collecting all the links pointed to from all the tweets, including source tweets and replies, and (2) by collecting the links that were manually annotated, as described in Section 5.1. With the collection of each of these links, we store the following data:

- The content of the link.
- The equivalent long URL for the link, given that many links on Twitter are shortened.
- The type of media (text, image, video, etc.).
- The title of the web page, except when it is image or video, where it is not available.

Chapter 6

Defining a Crowdsourcing Methodology for the Annotation of Rumourous Conversations

In this chapter we set forth the crowdsourcing methodology we have developed for the annotation of conversations around rumourous social media posts. We begin by justifying the need for a commercial crowdsourcing approach such as CrowdFlower¹ over other alternatives, and then delve into the settings of the crowdsourcing jobs, describing first how we disaggregate the tasks to facilitate the work, and detailing then the parameters we specified.

6.1 Why the Annotation was Crowdsourced

Having as a goal a large-scale annotation of rumourous conversations sampled from the dataset we have put together in PHEME, we studied different ways to perform the annotation. Since this a time-consuming task, and consequently expensive, we wanted to come up with a solution that would be economically affordable, efficient and reliable. We have considered different approaches for the annotation of the conversations:

1. **Recruit local volunteers to do the annotations.**
2. **Use a free crowdsourcing platform.**
3. **Use a commercial crowdsourcing platform.**

We carefully studied all three possibilities. First of all, we observed that we needed a large number of people to perform the annotation work, given that the work is time-consuming, and it can become cumbersome for a single person when spending many

¹<http://www.crowdfLOWER.com/>

hours on it and may result in low quality work. This requirement made it impossible to have people recruited locally to do the job. Hence, we needed a crowdsourcing platform to reach out to a larger community of users who could do small parts of the job, so we studied the viability of using free crowdsourcing platforms. One of the best-known solutions for free crowdsourcing is CrowdCrafting², which enables the submission of annotation work to be performed for free by others. We found two main issues with this alternative: (1) the users who do the jobs in free crowdsourcing platforms would like to get something back if they do not get paid, which can happen, for instance, with the annotation of health-related data that they can find beneficial even if only indirectly for themselves, and (2) we noticed that the completion of most of the jobs was very slow in this platform, occasionally taking up to 2 years to complete small batches of jobs. Since our annotation would probably not provide the users any non-economic benefit, and we would need to have the results in a reasonable time after submitting them, this alternative was not viable. It is also worth mentioning that while other partners within the PHEME consortium such as USH do have a platform and experience in crowdsourcing reports during unfolding events, this is not applicable to our scenario where we are seeking annotations for conversational aspects as observed in Twitter.

Hence, having found that the first two alternatives would not work for us, we ended up using a paid crowdsourcing platform so as to maximise speed [Procter et al., 2013a]. Crowdsourcing has been used extensively for the annotation of Twitter corpora in similar works for natural language processing [Finin et al., 2010, Paul et al., 2011]. After studying different alternatives, we chose CrowdFlower³ as it provides a flexible interface and has fewer restrictions than Amazon Mechanical Turk⁴.

6.2 Disaggregating Annotation Task into Microtasks

The annotation of an entire rumourous conversation can become time-consuming and cumbersome as it involves the annotation of all four features for all tweets in a conversation [Zubiaga et al., 2015]. As a first step we split the conversation into triples, where each triple consists of a tweet, which replies to the source tweet either directly or indirectly, its parent tweet (the tweet it replies directly to) and the source tweet (see Figure 6.1). If the tweet replies directly to the source tweet and no other previous tweet in the conversation then this is a tuple rather than a triple. Where the objective is to annotate the source tweet, this will appear on its own. Along with these tweets, we also show annotators the title assigned to the conversation during the rumour identification phase (see Chapter 5), which facilitates crowdsourced annotation of conversations by keeping in focus what the rumour is about.

To facilitate the task of the annotators further [Cheng et al., 2015], we narrowed down

²<http://crowdcrafting.org/>

³<http://www.crowdfLOWER.com/>

⁴<https://www.mturk.com/>

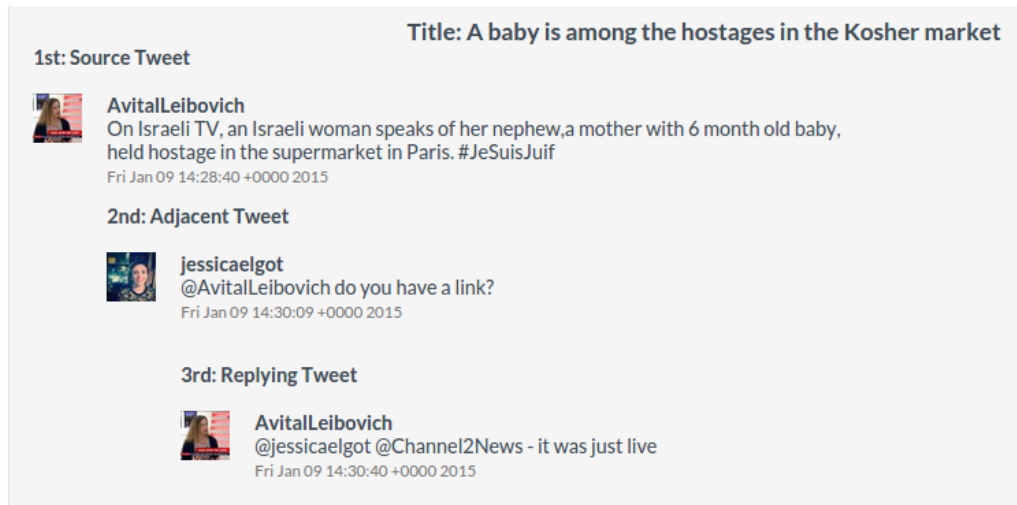


Figure 6.1: Example of a tweet triple as shown to the annotators.

the annotation unit to a single feature for each tweet triple, i.e., an annotator that accepts a microtask would be able to focus on a single feature (e.g. Response Type) without having to switch to other features. This can significantly speed up the process of annotating the same feature across triples or even different conversation threads. An alternative way of narrowing down the annotation task unit would be to ask each worker to annotate all the features for a single tweet. However, this would involve having to focus on different features, understanding the annotation guidelines for all of them at the same time, and requires more effort and concentration. Instead, our approach lets workers focus on a single feature, which makes the task guidelines easier to read and understand well. The disaggregation produced a total of 10 different microtasks that we then set up in the crowdsourcing platform. These 10 microtasks include 3 tasks for source tweets (annotation of each of support, certainty, evidentiality), 3 tasks for first-level replies (annotation of response type wrt the source tweet, certainty, evidentiality), and 4 for deep replies (annotation of response type wrt the source tweet, response type wrt the previous tweet, certainty, evidentiality). Each of these represent a separate job on the crowdsourcing platform.

6.3 Crowdsourcing Tasks Parameters

Our annotation units consist of either a triple/tuple of tweets or a single source tweet, annotated for a particular feature. For each annotation unit we collected annotations from at least 5 different workers. Each CrowdFlower job consists of 10 annotation units as described above. Thus this is the minimum an annotator commits to when accepting a job. We paid \$.15 for the annotation of each set of 10 units. In order to make sure that the annotators had a good command of the language in which the tweets are written, we re-

strict participants to relevant geographical areas. In our cases, we restrict the participants to those from the United States and the United Kingdom for English tweets, and to those from Germany and Switzerland for German tweets.

We performed an initial test on CrowdFlower to evaluate these parameters, which allowed further optimisation for the final crowdsourcing task. The initial tests helped us optimise the settings in the following two aspects. Firstly, we identified that having always 5 annotators (as was our initial configuration) was not optimal, as often more annotators were needed to reach agreement (defined below) in difficult cases. Thus, we enabled the *variable judgments mode* which allows us to have at least 5 annotators per unit, and occasionally more, up to a maximum number of annotators until a confidence value of 0.65 is reached. In most cases it was sufficient to set the maximum number of annotators to 7, apart from evidentiality where it was set to 10. Evidentiality is more challenging as one can assign 7 different values and more than one option can be picked, thus increasing the chance for a diverse set of annotations. Secondly, we noticed that some annotators were completing the task too fast, annotating a set of 10 units in a few seconds. To avoid this, we changed the settings to force the annotators to spend at least 60 seconds annotating sets of 10 source tweets, and at least 90 seconds annotating sets of 10 units of replying tweets.

To further guarantee the high quality of the annotations, we also created test questions for each annotation job. Test questions are sample tweets that we submitted with their associated annotations, and the annotators needed to match at least 70% of our ground truth annotations to quality for the job. This step in turn helped us get rid of underperformers that could harm the quality of the annotations.

The resulting settings have been used in subsequent crowdsourcing jobs, which we will describe in Chapter 8.

Chapter 7

An Annotation Scheme for Rumours

Having studied existing annotation schemes and their suitability for our purposes, we set out to develop a new annotation scheme adapted to the context of conversational threads around rumours in social media. This annotation scheme needs to be as generalisable as possible to different kinds of rumours that are discussed and disputed in social media, providing annotations that will enable the study of both linguistic aspects of the conversations, as well as sociological aspects that can be observed in the behaviour of participants.

To define this annotation scheme, we have followed an iterative process where it has been progressively tested and refined. First, we defined an initial annotation scheme that was based on the aforementioned schemes, which was then tested by assessing rumourous conversations extracted from Twitter. These rumourous conversations have the form of a thread, where a tweet starts the conversation, and subsequent tweets reply to that or other replying tweets, all of them having a parent tweet (see Figure 5.2 for an example of a rumourous conversation). This testing brought to light a set of strengths and weaknesses in this initial scheme, which was then refined in a new version. This new scheme was then tested again, in this case using also a crowdsourcing platform, to validate the changes. This chapter describes the steps of this process, showing the resulting annotation scheme. We will then follow in the next chapter by describing how we put it into practice with the creation of the annotated corpora that is being used in the PHEME project for the study of social media rumours.

7.1 Preliminary Annotation Scheme

Here, we take up the development of the annotation scheme from the latest version we presented in the earlier deliverable D2.1 [Zubiaga et al., 2014]. Then, after having tested an initial annotation scheme with two people on site, we identified certain characteristics for simplifying the annotation scheme and for making it simpler to understand and affordable to be used for manual annotation. These annotation tests helped us identify both the

strengths and the weaknesses of the initial annotation scheme. The test helped us find out that some of the features were suitable as they were, while others needed to be combined as they were adding redundant information and others needed to be slightly redefined.

The annotation scheme included, since the very beginning, two distinguishable parts in each rumourous conversation. These two parts include the source tweet, which is the one that starts a rumourous conversations, and replying tweets, which are all those that reply to that tweet. From the discussions after the first annotation tests, there was a strong agreement that the differentiation of two parts in a rumour was necessary. This is true especially when it comes to the source tweet, which is the one that introduces the rumour, and needs to be differentiated from the rest as replying tweets. However, there was a suggestion to redefine the other part, originally referred to as *spread and reactions*, which has now been renamed as *replying tweets*. Instead of annotating the conversation as a whole by asking each annotator to go through all the tweets in a conversation, which was rather complicated due to the need of aggregating all the responses and providing a single annotation, the new proposal was to annotate each response tweet separately. Borrowing from conversational analysis and the concept of *adjacency pairs* and *turn taking*, we argue that each response tweet should be understood as a posting that is paired with some previous tweet – i.e., the former is either a retweet, a reply to or otherwise mentions the author of the latter. From this, we conclude that each tweet subsequent to the source tweet should be annotated in the context of the tweet to which it is paired – that is it is annotated for how it can be seen to stand in relation to that particular preceding tweet. Therefore, we relabel the two parts of the annotation scheme from originally referred to as *message crafting* and *spread and reactions* to the new labels defined as *source tweet* and *replying tweets*.

For these two newly defined parts of the annotation scheme, we then defined the features that were deemed relevant and discarded the rest. The resulting annotation scheme is shown in Figures 7.1 (for source tweets) and 7.2 (for replying tweets). We consider this as the preliminary annotation scheme at this point, which we have since tested and revised further to come up with the final annotation scheme that we will describe later in this deliverable.

We will now describe the annotation tests conducted with this preliminary annotation scheme, and how this helped us to come up with the final version.

7.2 Validation and Revision of the Preliminary Annotation Scheme

While we had an earlier version of the annotation scheme tested by two people on site, we wanted to further test this preliminary annotation scheme. This time, we enlisted the help of two experienced PhD students in Applied Linguistics, who had prior experience in Conversation Analysis. One of them had also long term experience with Twitter as a

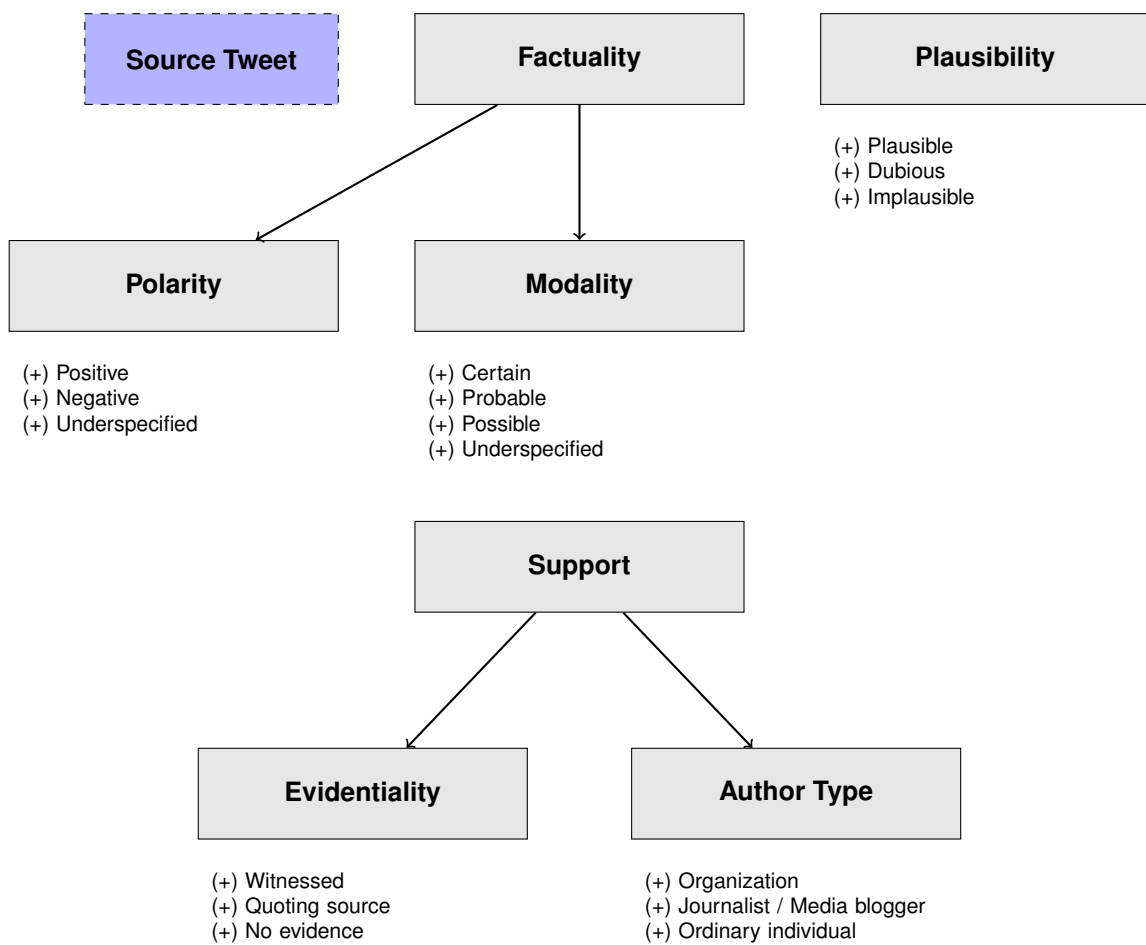


Figure 7.1: Annotation scheme for source tweets that initiate rumors

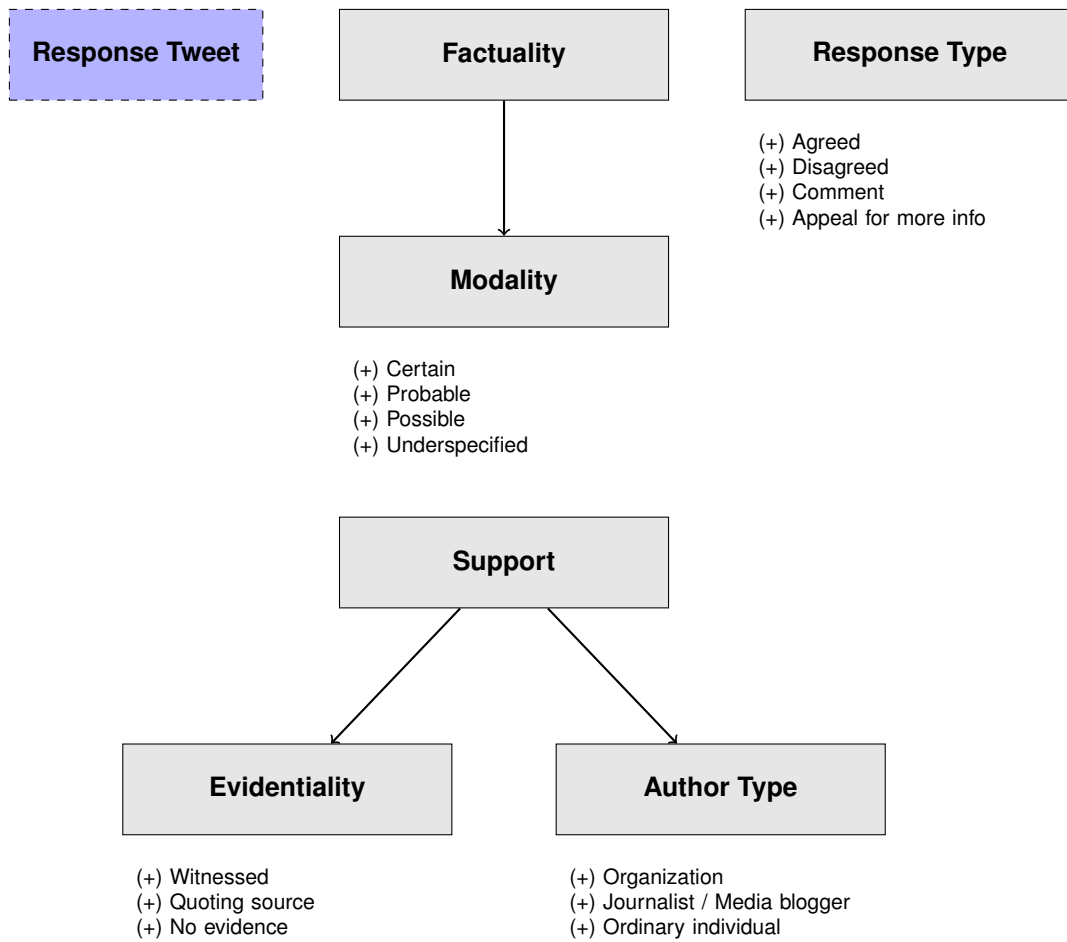


Figure 7.2: Annotation scheme for tweets responding to the initial rumourous tweets, as well as subsequent responses

user. We sought people with this expertise for further help revising the scheme. These two students spent one hour each with us, thinking aloud while they were annotating rumourous conversations. Being still a small-scale annotation, we did not make use of the crowdsourcing platform, but we continued using the annotation tool we developed to this end. The crowdsourcing platform is instead used in subsequent steps after we revised this preliminary annotation scheme.

Thanks to the feedback we got from previous annotation tests, we could update the annotation tool accordingly. Figure 7.3 shows the revised version of the annotation tool. The main update with respect to the previous version of the tool used is that the features to be annotated are presented one by one (instead of all at once as before), and that the questions are much more descriptive than before, so it should be easier for the participants to remember what e.g. “plausibility” means. For instance, instead of asking them to annotate “modality”, we ask them “Is the author confident about their statement?”. The responses are also more descriptive now, e.g., for the question above, we show the following possible answers: a) Yes, they are entirely confident, b) They are slightly unsure, c) They are not very sure, and d) It’s unclear.

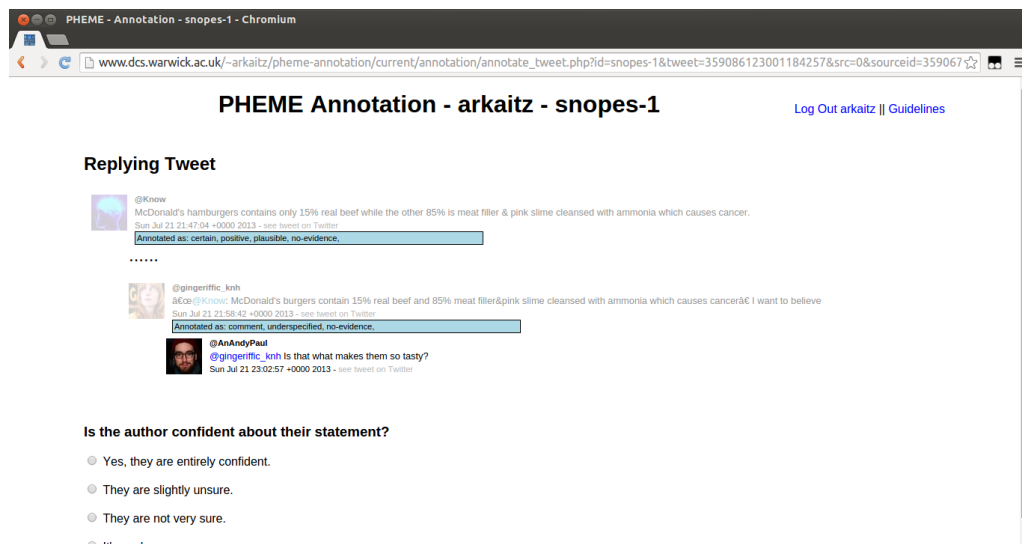


Figure 7.3: Screenshot of the revised tool for annotation of rumourous conversations.

In these new tests we could identify numerous improvements needed in the annotation scheme, in comparison with those we found out with the earlier tests. The most important improvement, and probably the main issue with the previous version, is that the participants felt much more comfortable while doing the annotations from the very first tweet (they needed about 10 tweets in the previous tests to get used to the scheme). The fact that features are now shown one by one, and that they are associated with descriptive questions and answers, helped them to get familiar with the tool and the scheme much faster. The fact that they did not ask about the meaning of some feature at any point shows a

significant improvement in this sense. This helps both reduce the amount of information needed to include in the guidelines, as well as the time needed to learn how to annotate.

Besides the need for disaggregating the whole annotation task into smaller microtasks, we identified additional improvements to be applied to the annotation scheme. One of the major changes we did at this point in the annotation scheme was to combine both schemes for source tweets and replies into a single one. We did this because we noticed that most of the features were equivalent for sources and replies, except for the support. Response type, annotated only for replies, has the same values as in the previous version: agreed, disagreed, appeal for more information, or comment. However, it is worth mentioning that it is annotated twice, as we felt the need to annotate the type of response that a tweet represents with respect to the previous tweet as well as with respect to the source of the thread. What was previously referred to as polarity for source tweets, it is now called support, where its possible values have also been renamed, while the meaning is similar: positive changes to supporting, and negative to denying.

We also identified the need for expanding the types of evidence that could be annotated. While we only had three before, this was often insufficient, as the annotators suggested, and therefore we revised evidentiality to include seven different options: (1) first-hand experience, (2) pointing to URL with evidence, (3) quotation of a person or organisation, (4) attachment of a picture as a proof of evidence, (5) quotation of an unverifiable source, (6) employment of reasoning, and (7) lack of evidence. This expanded list of values for evidentiality would enable us to further specify the type of evidence given in a tweet, especially to differentiate all different types of sources that can be quoted, which was not distinguished before.

Besides the changes above, we also needed to make the task a bit simpler to the extent possible, and noticed that we could take out two of the features from the original annotation schemes. On one hand, we noticed that rumours, as they are widely spread due to the uncertainty they produce, tend to be plausible by definition. We realised that most of the tweets were being annotated as being plausible, and therefore it was not useful to annotate for plausibility. On the other hand, it was becoming very difficult for annotators to annotate the author type. This occurred because the annotators are focussing their annotation work on the content of tweets, and moving then to an annotation related to the author of the tweet was difficult. It was also challenging to do this annotation as it requires looking at many different factors that each user has. Due to this, we opted for automatically inferring the author types from their metadata, as this can be possible for author type codeframe we defined in Section 2.3.

The resulting annotation scheme is shown in the revised scheme in Figure 7.4.

We define the features included in this revised annotation scheme as follows:

Support: Support is only annotated for source tweets. It defines if the message in the source tweet is conveyed as a statement that supports or denies the rumour. It is hence different from the rumour's truth value, and intends to reflect what the tweet suggests is author's view towards the rumour's veracity. The support given by the author of a

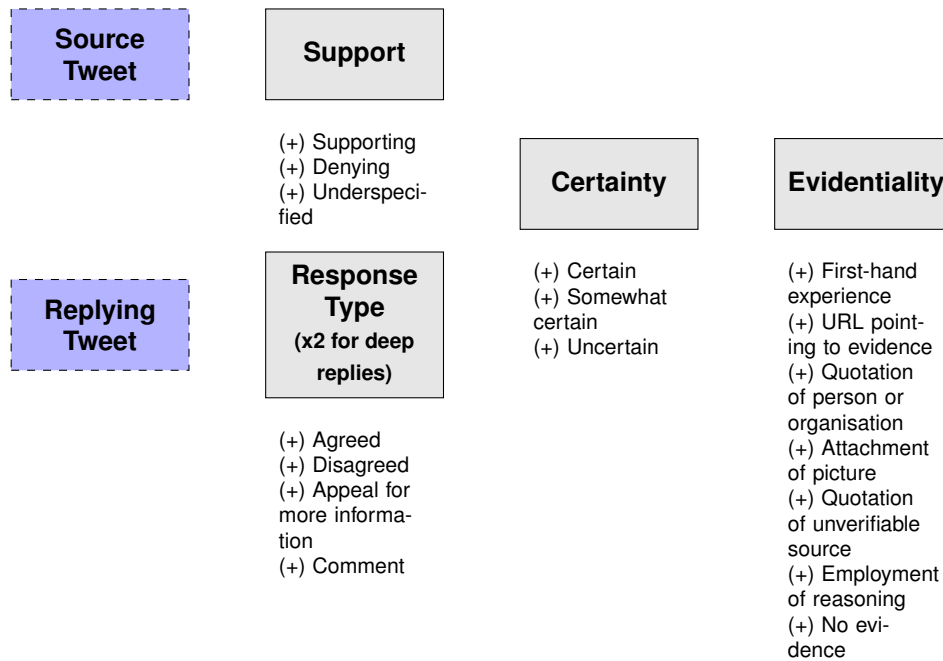


Figure 7.4: Annotation scheme for rumourous social media conversations.

tweet can be deemed as: (1) supporting the rumour, (2) denying it, or (3) underspecified, when the author’s view is unclear. This feature is related to the “Polarity” feature in the factuality scheme by Saurí et al. [Saurí and Pustejovsky, 2009].

Response Type: Response type is used to designate support for the replying tweets. Given a source tweet that introduces a rumourous story, other users can reply to the author, leaning for instance in favour or against the statement. Some replies can be very helpful to determine the veracity of the rumour, and thus we annotate the type of reply with one of the following four values: (1) *agreed*, when the author of the reply supports the statement they are replying to, (2) *disagreeing*, when they deny it, (3) *appeal for more information*, when they ask for additional evidence to back up the original statement, or (4) *comment*, when the author of the reply makes their own comment without adding anything to the veracity of the story. Note that the response type is annotated twice for deep replies, i.e., tweets that are not directly replying to the source tweet. In these cases, the response type is annotated for a tweet determining two different aspects: (i) how the tweet is replying with respect to the rumour in the source tweet, and (ii) how the tweet is replying to the parent tweet, the one it is directly replying to. This double annotation allows us to better analyse the way conversations flow, and how opinions evolve with respect to veracity. The inclusion of this feature in the annotation scheme was inspired by [Procter et al., 2013b], who originally introduced these four types of responses for rumours.

Certainty: Certainty measures the degree of confidence expressed by the author when posting a statement in the context of a rumour and applies to both source tweets and replies. The author can express different degrees of certainty when posting a tweet, from

being 100% certain, to considering it as a dubious or unlikely occurrence. Note that the value annotated for either support or response type has no effect on the annotation of certainty, and thus it is coded regardless of the statement supporting or denying the rumour. The values for certainty include: (1) *certain*, when the author is fully confident or the author is not showing any kind of doubt, (2) *somewhat certain*, when they are not fully confident, and (3) *uncertain*, when the author is clearly unsure. This feature and the possible values were inspired by [Saurí and Pustejovsky, 2009], who referred to it as “modality” when annotating the factuality of news headlines.

Evidentiality: Evidentiality determines the type of evidence (if any) provided by an author and applies to both source tweets and replying tweets. It is important to note that the evidence has to be directly related to the rumour being discussed in the conversation, and any other kind of evidence that is irrelevant in that context should not be annotated here. Evidentiality can have the following values: (1) *first-hand experience*, when the author claims to have witnessed events associated with the rumour (2) *attachment of a URL* pointing to evidence, (3) *quotation* of a person or organisation, when an accessible source is being quoted as a source of evidence, (4) *attachment of a picture*, (5) quotation of an *unverifiable source*, when the source being mentioned is not accessible, such as “my friend said that...”, (6) *employment of reasoning*, when the author explains the reasoning behind their view, and (7) *lack of evidence*, when none of the other types of evidence is given in the tweet. Contrary to the rest of the features, more than one value can be picked for evidentiality, except when “lack of evidence” is selected. Hence, we cater for the fact that a tweet can provide more than one type of evidence, e.g. quoting a news organisation while also attaching a picture that provides evidence.

7.3 Final Validation and Revision of the Revised Annotation Scheme

The revised annotation scheme was then further tested, in this case by making use of our real annotation scenario through crowdsourcing. We sampled a small subset of 8 threads, and tested the annotation scheme using crowdsourcing.

7.3.1 Dataset Sampling for Testing the Scheme

To validate and assess the viability of crowdsourcing annotations using our scheme, we sampled 8 different source tweets and their associated conversations from the 784 rumours identified for 3 of the events. These were the events that we had readily available at the time. This includes 4 source tweets for the Ferguson unrest, and 2 source tweets each for the Ottawa shootings and the rumourous story of Essien having contracted Ebola (Table 7.1 shows the number of source tweets and replies included in each case).

Event	Src. tweets	1st rep.	2nd rep.
Ferguson unrest	4	63	58
Ottawa shootings	2	20	35
Ebola	2	22	10
TOTAL	8	105	103

Table 7.1: Tweets sampled for annotation.

7.3.2 Validation through Crowdsourcing and Reference Annotations

This sample of 8 threads including 216 tweets was submitted to CrowdFlower for crowdsourced annotation. Through these tests, we collected the crowdsourced manual annotations for all features associated with the 216 tweets. This amounts to the annotation of 4,974 units (tweet triple+feature combination), and was performed by 98 different contributors. The final set of annotations was obtained by combining annotations by all workers through majority voting for each annotation unit. The cost for the annotation of all 8 threads amounted to \$102.78.

Having at least 5 annotators per tweet-feature pair, we could compute the agreement of annotators with one another, and measure the quality to some extent. However, for further testing and validation, we also wanted to measure how accurate they were compared to a ground truth that we would establish. In order to have a set of reference annotations to compare the crowdsourced annotations against, the whole annotation task was also performed by one member of the UWAR partner, which we use as a reference annotation (REF). A second annotator, a member of the SWI partner, annotated also one third of the whole (REF2). This allows us to measure three things: (1) agreement of crowdsourced annotators with one another, (2) agreement between the two reference annotations (REF and REF2), and (3) agreement between the crowdsourced annotators and the reference annotations.

7.3.3 Analysis of of the Crowdsourced Annotation

In order to report inter-annotator agreements, we rely on the percent of overlap between annotators, as the ratio of annotations that they agreed upon. Table 7.2 summarises the agreement values between different annotations. The inter-annotator agreement between REF and REF2 was 78.57% measured as the overlap. This serves as a reference to assess the performance of the crowdsourced annotations in subsequent steps. When we compare the decisions of each of the annotators against the majority vote, we observe an overall inter-annotator agreement of 60.2%. When we compare the majority vote against our reference annotations, REF, they achieved an overall agreement of 68.84%. While this agreement is somewhat lower than the 78.57% agreement between REF and REF2, it is only worse when annotating for “certainty”, as we will show later. This also represents

a significant increase from earlier crowdsourcing tests performed before revising the settings, where the annotators achieved a lower agreement rate of 62.5%. When breaking down the agreement rate for each of the features (see Table 7.3), we see that the agreement values range from 58.17% for certainty in reply tweets, to 100% for support in source tweets. The agreement rates are significantly higher for source tweets, given that the annotation is easier as there is only the need to look at one tweet, instead of tuples/triples. This analysis also allows us to compare the agreement by feature between the crowdsourced annotations (CS) and REF, as well as between REF and REF2. We observe that agreements are comparable in most cases, except for the agreement on certainty, which is significantly higher between REF and REF2. The latter represents the major concern here, where the crowdsourcing annotators performed worse, which we explain later.

	CS	REF
CS	60.2%	68.84%
REF	-	78.57%

Table 7.2: Inter-annotator agreement values between different annotators.

Source tweets			
	Support	Certainty	Evident.
CS vs REF	100%	87.5%	87.5%
REF vs REF2	100%	62.5%	87.5%
Replying tweets			
	Resp. type	Certainty	Evident.
CS vs REF	70.42%	58.17%	74.52%
REF vs REF2	71.82%	87.14%	78.89%

Table 7.3: Inter-annotator agreement by feature.

In more detail, Table 7.4 shows the distribution of annotated categories, as well as the agreement rates for each feature when compared to the reference annotations, REF. Looking at the agreement rates, annotators agreed substantially with the reference annotations for source tweets (100% agreement). For replying tweets, as discussed above, the depth of the conversation and the additional context lead to lower agreement rates, especially for some of the categories. The agreement rates are above 60% for the most frequent types of values, including response types that are "comments" (67.69%), authors that are "certain" (60.22%), and tweets with "no evidence" (85.37%). The agreement is lower for the other annotations, which appear less frequently. This certainly proves that the annotation of replies is harder than the annotation of source tweets, as the conversation gets deeper and occasionally deviates from the topic discussed in the source tweet. One of the cases with a low agreement rate is when the evidence provided is "reasoning". This shows the need to emphasise even more in subsequent crowdsourcing tasks the way this type of evidence

Source tweets					
Support		Certainty		Evidentiality	
% of times	agreem.	% of times	agreem.	% of times	agreem.
supporting (100%)	100%	certain (75%)	100%	no evidence (37.5%)	100%
denying, underspecified (0%)	–	somewhat certain (12.5%)	100%	author quoted (37.5%)	100%
	–	uncertain (75%)	100%	picture attached (25%)	50%
				URL given, unverifiable source, witnessed, reasoning (0%)	–
					–
Replying tweets					
Response type		Certainty		Evidentiality	
% of times	agreem.	% of times	agreem.	% of times	agreem.
comment (66.56%)	67.69%	certain (54.33%)	60.22%	no evidence (79.81%)	85.37%
disagreed (15.43%)	53.70%	somewhat certain (25.96%)	40%	reasoning (9.62%)	29.17%
agreed (10.61%)	50%	uncertain (19.71%)	41.18%	author quoted (3.37%)	62.5%
appeal for more info (7.40%)	33.33%			URL given (3.37%)	50%
				picture attached (2.89%)	33.33%
				witnessed (0.48%)	0%
				unverifiable source (0.48%)	0%

Table 7.4: Distribution of annotations: percent of times that each category was picked, and the agreement with respect to our reference annotations (CS vs REF).

should be annotated, by remarking that the reasoning that is being given in a tweet must be related to the rumourous story and not another type of reasoning.

When we look at the distribution of values the annotators chose, we observe an imbalance in most cases. For response type, we see that as many as 66.5% of the replies are comments, which shows that only the remainder 33.5% provide any information that adds something to the veracity of the story. The evidentiality is even more skewed towards tweets that provide no evidence at all, which amount to 85.4% of the cases. Both the abundance of comments, and the dearth of evidence, emphasise the need for carefully analysing these conversations when building machine learning tools to pick out content that is useful to determine the veracity of rumourous stories. The certainty feature is slightly better distributed, but still skewed towards more than 54% cases of certain statements; this could be due to the fact that many users do not express uncertainty in short, written texts even when they are not 100% sure.

To better understand how the different features that have been annotated fit together, we investigated the combinations of values selected for the replying tweets. Interestingly, we observe that among the replying tweets annotated as comments as many as 80.3% were annotated as having no evidence, and 47.5% were annotated as being certain. Given that comments do not add anything to the veracity of the rumour, it is to be expected that there would be no evidence. We also investigated several cases to understand how certainty was being annotated for comments; we observed that different degrees of certainty were being assigned to comments where certainty can hardly be determined as it does not seem to apply, e.g., in the tweet “My heart goes out to his family”. This also helped us understand the low agreement rate between CS and REF for certainty, which may drop due to the comments with an unclear value of certainty. For these two reasons, together with the fact that comments represent tweets that do not add anything to the veracity of the story, we consider revising the annotation scheme so that these two features should not be annotated

for comments. This, in turn, reduces significantly the cost of running the crowdsourcing tasks, given that for as many as 66.5% replying tweets that represent comments, we would avoid the need for two annotation tasks.

7.4 Final Annotation Scheme

After the final tests by combining crowdsourced annotations and our expertise, the validity of the annotation scheme was largely corroborated and so we were ready to move on with a larger scale annotation of rumourous conversations. The final annotation scheme, which is slightly revised from the previous version, includes two changes: (1) the evidentiality and certainty will not be annotated for replies deemed comments, which was found unnecessary and saves a significant amount of work and money, and (2) the certainty will now include an additional value, underspecified, for the cases where the degree of certainty of the author cannot be determined because of the brevity of a tweet or the lack of detail. The resulting annotation scheme, which is the final and validated annotation scheme used in PHEME is shown in Figure 7.5.

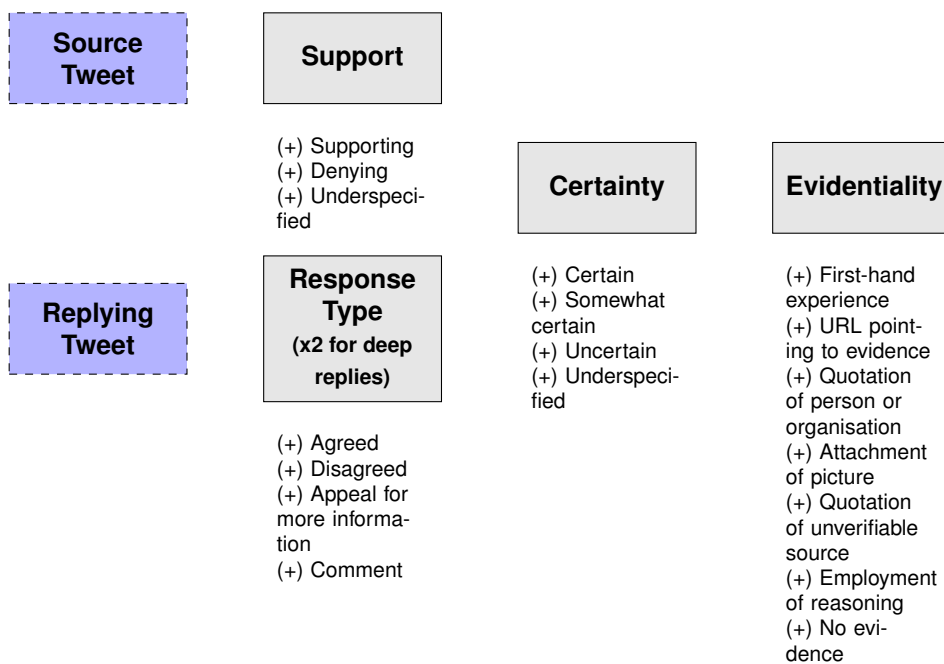


Figure 7.5: Annotation scheme for rumourous social media conversations.

Chapter 8

Dataset Annotation

Having come up with a dataset of rumourous conversations, a validated annotation scheme, as well as a crowdsourcing methodology, our next step was to perform a large-scale annotation of conversations, so that we can study rumours in social media. In this chapter, we describe the process we followed to collect the annotations through crowdsourcing, including both dataset sampling and the annotation itself. Finally, we also summarise the outcome of the crowdsourced annotation work.

8.1 Dataset Sampling for Crowdsourced Annotation

The dataset we used here contains initially 2,695 rumourous threads: 2,460 in English, 198 in German, and 37 in French. Given that the PHEME consortium has partners who can fluently speak, and have previously developed linguistic analysis tools for English and German, we focused our annotation efforts on these two languages. This gave us 2,658 rumourous threads to sample from, after removing the French threads.

Since the annotation of all the rumourous threads would be both time-consuming and economically unaffordable, we sampled a subset of the original set so as to obtain a representative sample that will allow us to perform the analysis. The amount of data to be sampled has been determined by both the data we would need for a compelling analysis, as well as the annotation work that is economically affordable. Thanks to the involvement and contribution from USFD and USAAR, together with UWAR, each contributing with \$500 to fund the crowdsourcing jobs, the sampling has been performed with a total budget of \$1,500.

Since threads are comprised of tweets, and the number of tweets can vary across threads, we first wanted to estimate the approximate cost of annotating a tweet, so we could then estimate the number of threads that we could afford to annotate. From our earlier tests using the crowdsourcing platform, the average cost of annotating a tweet for the different features was approximately \$0.33. With this in mind, we prepared the data

sampling process.

On the other hand, we wanted to make the annotation task as effective as possible, and therefore we wanted to avoid annotating noisy or unnecessary tweets. Hence we filtered the dataset by taking out tweets that met at least one of the following characteristics:

- The replying tweet is in a different language from that of the source. This may lead to the annotation of tweets in other languages, especially languages that the annotators might not be fluent in, and we would be wasting money. Hence, we only consider replying tweets in English or German for source tweets in the same language, as well as tweets labelled as “undeterminable”, which are usually tweets with very little textual context or perhaps only a link, where the language cannot be determined from the tweet itself.
- The replying tweet is a manual retweet of a previous tweet in the thread. We did this by checking the degree of duplication between a tweet and its replying tweets. We removed tweets that did not have at least a Levenstein difference of 5 [Navarro, 2001]. This was to allow for the changes that users retweeting sometimes make to add comments or to ensure that the resulting retweet still fits within the 140-character format.

In the aforementioned cases, we skipped the manual annotation as it would not be suitable, but we still kept them in the thread for the analysis, in this case with no annotation.

Besides defining these constraints for optimisation, the data sampling had to be fairly performed so that the distribution of rumours was still representative of the whole. We especially wanted to be careful about selecting tweets from all 9 events in our dataset, across different stories, and including threads that were posted both before or after turnaround tweets, as well as the turnarounds themselves. We developed a script for data sampling that fulfils our requirements.

The data sampling led to a subset with 330 threads, 297 in English and 33 in German. Having removed unwanted replying tweets from these threads, the resulting dataset contains 4,842 tweets: 4,560 tweets in English, and 282 tweets in German. This dataset was submitted to the crowdsourcing platform for annotation.

8.2 Annotation of Rumourous Conversations

The crowdsourced annotation work was again conducted using CrowdFlower, with the same parameters as specified in 6 and also used for the initial testing and validation of the annotation scheme. The annotation jobs were mainly split into two, one for English and one for German, as they require slightly different settings in terms of geographical restrictions of participants. Tweets for the 9 different events were put together within the

same annotation jobs for each language, so that the annotation work is less cumbersome than for a single event with repetitive stories. Given that the dataset for English is quite large, the tweets to be annotated were submitted to CrowdFlower progressively, making sure at all times that the work was being performed properly. The whole annotation process including both languages took approximately three weeks to complete.

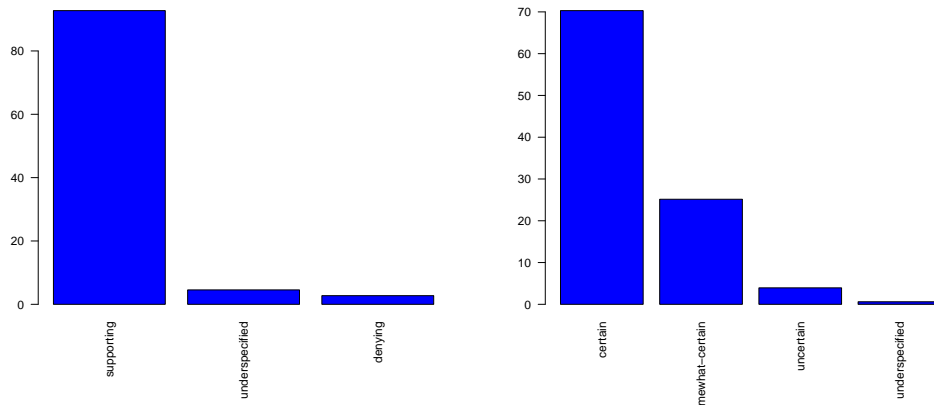
8.3 Outcome of the Crowdsourced Annotation

The whole annotation work consisted of 68,247 judgments on 4,842 tweets performed by 233 different annotators. We combine all judgments for each tweet feature pair through majority vote. To quantify the difficulty of the task, we measure the inter-annotator agreement values by comparing each judgment with the majority vote. Overall, the annotators achieved an agreement rate of 62.3%, which is distributed differently across different tweet types and features. Table 8.1 shows how the agreement rates are distributed for source tweets and replies when annotating for support, certainty, and evidence. As we expected, this shows that the annotators found it easier to annotate source tweets, as we have also found before that they are less ambiguous and requiring less context for understanding, as the tweet alone usually makes sense. The agreements are somewhat lower for replying tweets. When we compare the different features, we observe that support is the easiest to annotate for source tweets, but very similar to certainty overall. Evidentiality is slightly more difficult to annotate, most probably because of the large number of different values that the annotators can choose.

	Support	Certainty	Evidence
Source tweets	81.1%	68.8%	74.9%
Replies	62.2%	59.8%	58.3%
Overall	63.7%	61.1%	60.8%

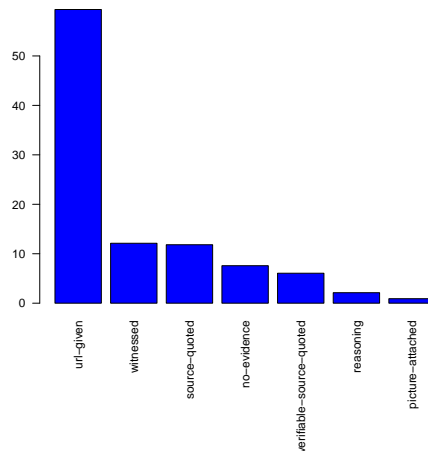
Table 8.1: Inter-annotator agreement values for different features and tweet types.

We visualise the aggregated annotations for the 330 threads in Figures 8.1 and 8.2. Figure 8.1 shows the annotations for source tweets, where we can observe that a majority of source tweets support the rumour in question, the author is certain, and evidence is provided by pointing to an external article. Figure 8.2 shows the annotations for replying tweets, where the majority of replies are comments which do not add anything to the veracity of the story, the author is certain, and no evidence is provided. At a first glance, this shows a big difference between source tweets and replying tweets, where the latter need to be carefully analysed to be able to find the useful opinions and evidence contributing towards the clarification of the veracity of the rumour.



(a) Support

(b) Certainty



(c) Evidentiality

Figure 8.1: Distribution of annotations for source tweets.

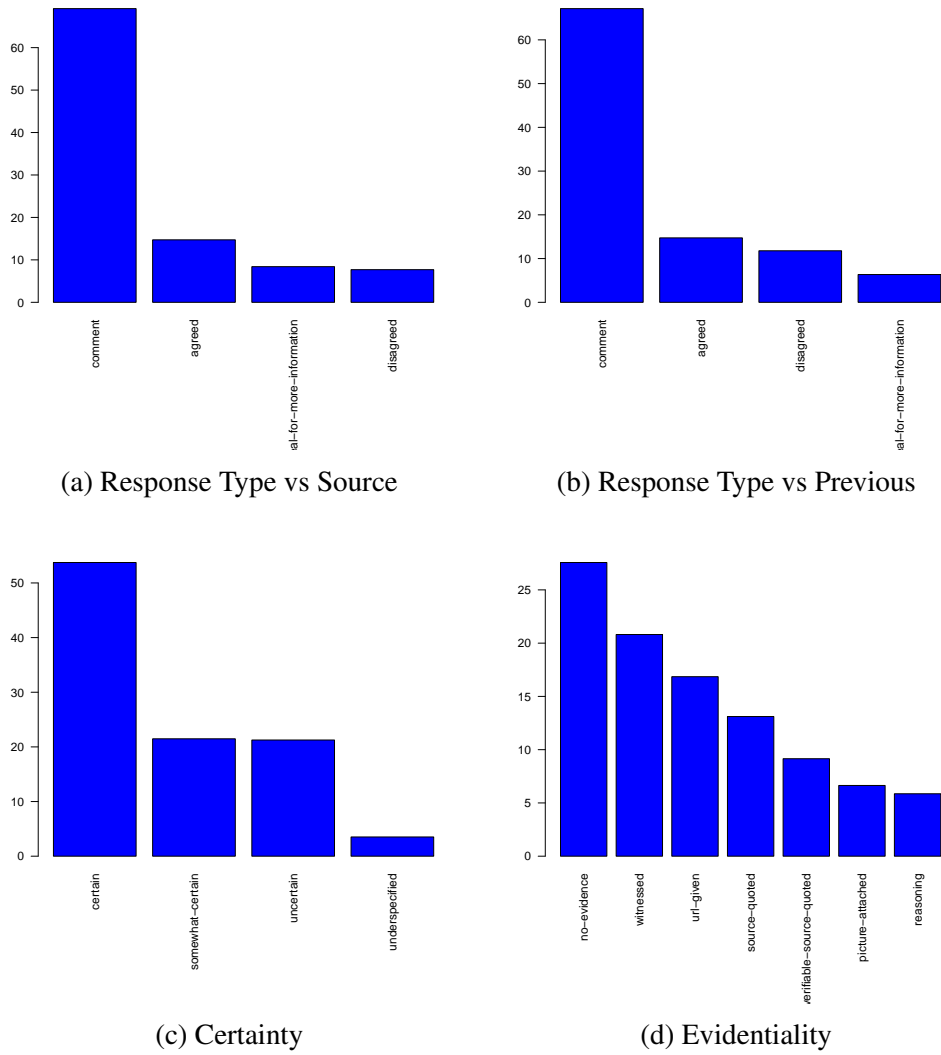


Figure 8.2: Distribution of annotations for replying tweets.

Chapter 9

Extending the Dataset with Rumour and Actor Types

This chapter describes the final adjustments to enable the qualitative analysis of rumourous conversations in social media, looking at the interactions of different actor types across different types of rumours, media, and languages. We describe the preprocessing step we conducted for inferring the rumour types and actor types to enable the qualitative analysis from our dataset.

9.1 Inferring Rumour Types and Actor Types

We make use of the annotations we gathered through crowdsourcing, as well as additional information sources we describe below, to determine the actor types involved in the conversations, as well as to distinguish the different types of rumours. This categorisation allows us to perform the qualitative analysis of rumourous conversations in social media.

9.1.1 Determining Rumour Types

As described in Section 2.2, we categorise rumour types by two different factors: accuracy and acceptability. The categorisation by accuracy was performed in the manual annotation process performed by SWI, which led to the following distribution for the 330 threads in our sample: 159 are true, 68 are false, and 103 remain unverified. On the other hand, the categorisation of rumours by acceptability includes three different types, which we infer from our sample as follows:

1. **Speculation**, which we consider as the early reports that lack supporting evidence. Hence, source tweets annotated with no evidence will be considered as speculation here. This categorisation led to the identification of 25 threads as being speculative.

2. **Controversy**, which we detect by looking at the percentage of replying tweets that either disagree or appeal for more information.
3. **Agreement**, being the opposite of controversy, is computed as the percentage of replying tweets that are not disagreements or appeals for more information.

For computing the percentage of controversy or agreement, we measure the controversy level as the percentage of replies which are disagreeing or appeal for more information. Rather than establishing a threshold here to determine what is controversy and what is not, we create a ranking of threads by their level of controversy from 0% to 100%, where those with the lowest level of controversy can be deemed cases of agreement.

9.1.2 Categorising Users by Actor Type

As described in Section 2.3, we categorise actors by type based on three different factors: (1) whether they are verified users or not, (2) their follow ratio, and (3) whether they are journalists or news organisations, or not.

The fact of a user being or not verified can be directly inferred from a tweet's metadata. By looking at this value from a tweet's metadata, we found that in our sample, only 171 users are verified and the remainder 3,381 are non-verified users. The small subset of users with this condition shows the importance of such a feature.

To compute the follow ratio of each user, we compute the difference in terms of orders of magnitude between the number of users that follow them and the number of users they follow. This is computed by using the following equation:

$$\log_{10}(\#followers/\#following) \quad (9.1)$$

Figure 9.1 shows the distribution of these follow ratios for the users in our sampled dataset. It can be seen that the most frequent ratio is 0 (users follow and are followed by an amount of users in the same order of magnitude), and there are fewer users who have higher follow ratios (1 to 7, i.e., with more followers than followees) or even lower (-1, i.e., following more people than follow them). Again, we show that the follow ratio helps us distinguish a few outstanding users, which is of help for our qualitative analysis.

Finally, we want to distinguish journalists and news organisations from the rest of the users. While there is no perfect solution to infer this from a tweet's metadata, we have created lists of journalists and news organisations from the Web in order to achieve this categorisation of users. For the list of Twitter accounts for news organisations, we relied on the manually curated directory at Muckrack¹, and for the list of journalists, we put together numerous lists created by reputable journalists and news media on Twitter^{2 3}

¹<http://muckrack.com/media-outlets>

²<https://twitter.com/bbcbreaking/lists/news-sources>

³<https://twitter.com/arjunkharpal/lists/news-organisations>

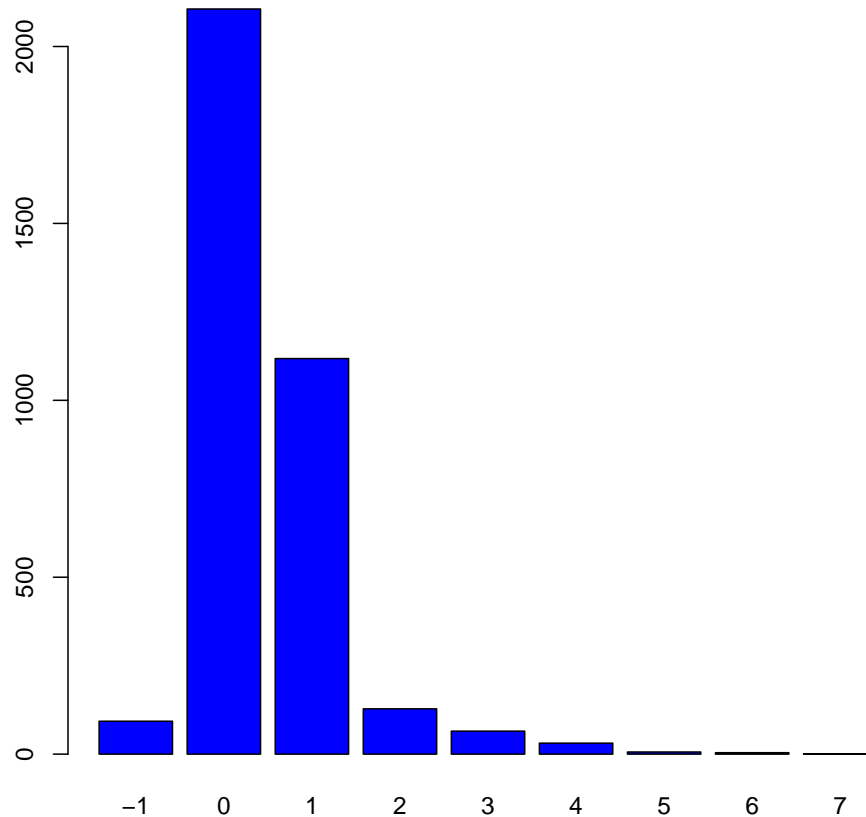


Figure 9.1: Distribution of follow ratios for users in our sampled dataset.

⁴ ⁵ ⁶ ⁷ ⁸. Using these lists, we could come up with extensive lists including 513 news organisations, and 24,748 journalists.

When matching this list of users to our dataset, we identified that 89 users are journalists or news organisations, while 3,463 are not. Again, the subset of professionals in journalism is very small here, which makes the analysis of this type of users more important for our study.

⁴<https://twitter.com/ftistratcomm/au/lists/news-organisations-au>

⁵https://twitter.com/jj_bryant/lists/news-organisations

⁶<https://twitter.com/mashable/lists/news>

⁷<https://twitter.com/nytnational/lists/news-organizations>

⁸https://twitter.com/sam_ikin/lists/news-organisations

Chapter 10

Extending the Annotation Scheme

As already indicated in the earlier sections of this document, the other aim has been to build upon the annotation scheme outlined in Chapter 7 by exploiting existing work in the areas of conversation analysis and ethnomethodology. This approach has offered us the possibility of: a) uncovering richer ways to annotate how Twitter feeds and rumours around particular topics unfold across extended threads and sequences of action; and b) grounding annotations in a way that will enhance their capacity to capture features of people's own situated reasoning. In this section we look a little more closely at what pursuing this approach to grounding annotations looks like. Central to the proposition is the notion that streams of Twitter feeds around the same topic can be conceptualized in some way as conversations. However, as one begins to explore the ways in which such a suggestion might be justified, one begins to also realise that there are ways in which tweeting stands as an independent phenomenon (separate from regular conversation analysis) that needs to be understood on its own terms. Thus we shall be arguing here that, whilst conversation analytic approaches serve well as a point of departure, differences between spoken conversation and the organisational character of tweet-based interaction make it necessary to respecify the approach more precisely as 'microblog analysis' in order to steer around the potential dangers of missing the lived character of how people reason about tweeting as an activity in its own right.

10.1 Moving towards annotation grounded in microblog analysis

Over the course of this section we will be looking at the conceptualization of tweets as conversations a little more closely and exploring the ways in which similarities do exist and the ways in which tweeting may be seen to present discrete phenomena that cannot easily be subsumed within conventional approaches to conversation analysis.

10.1.1 The turn-taking mechanism

At the very heart of conversation analysis, as laid out by [Sacks et al., 1974], is the observation that talk is organised such that only one speaker speaks at once. This is seen as a fundamental premise of social order because any other system would frequently render talk completely ineffectual. On the basis of this, and probing just how it could be that this is systematically provided for in interaction, Sacks et al. elaborated what they called the 'turn-taking mechanism'. It contains some primary features that together serve to underpin most other kinds of conversational phenomena. So there are: speakers (recognizable individuals who produce utterances); speakers who talk first, and other speakers who may also talk as a conversation unfolds; mechanisms whereby a current speaker may select who talks next; and mechanisms whereby speakers may select themselves to be the next person to produce an utterance.

On the basis of a number of years of close examination of conversational data Sacks and his colleagues assembled a highly robust model of turn-taking in conversation that can be seen to have a number of key strengths. One of the most important aspects of all is that the proposed model is able to be simultaneously 'context-free' but also exceptionally 'context-sensitive'. So you can dip into whomsoever, wheresoever and find the same system in play, with the same key operational characteristics. At the same time, the system can be endlessly adapted to meet the particularities of local need without having to step outside of the system itself ([Sacks et al., 1974]: 700).

The more specific observations Sacks et al make regarding the actual workings of the turn-taking system for conversation are of varying degrees of applicability to our own interest in tweet exchange in Twitter. The four primary observations are (see [Sacks et al., 1974]: 700-701):

1. "Speaker-change recurs, or at least occurs".
2. "Overwhelmingly, one party talks at a time".
3. "Occurrences of more than one speaker at a time are common, but brief".
4. "Transitions (from one turn to a next) with no gap and no overlap are common. Together with transitions characterized by slight gap or slight overlap, they make up the vast majority of transitions".

These first four characteristics of the turn-taking system in conversation are, in large part, managed within Twitter by its technical configuration so function more at the level of given constraints than situated accomplishments. Thus speaker (or tweeter) change is a direct function of who is being followed, the frequency with which they tweet, and the presence of other factors such as promoted tweets. It is conceivable that someone might follow just one other party in which case tweeter change would be rare. However, promoted tweets usually result in some extraneous tweets appearing on anyone's timeline

throughout the day. More importantly, in view of the fact that even in 2012 the average number of people being followed for Twitter users was 102¹ and Twitter has expanded since then, tweeter change is a characteristic of most people's timelines and, for the larger part, two or more tweets concurrently by the same person is infrequent though it certainly occurs. Similarly, it is a feature of Twitter that the timeline is organised in independent tweets that do not appear in a simultaneous and overlaid fashion and that do not overlap. So each turn is tightly independent and consecutive. Temporal gaps between the appearance of one tweet and another do routinely occur, however, though they do not manifest themselves as 'gaps' in the timeline but rather as delays in updates. Once again the extent to which delays in updates occur is tightly bound up with the number of people being followed and the frequency with which they tweet. Broadly speaking, though, temporal disjuncture between tweets can be considered to be a routine feature, making gaps in interaction an unremarkable feature of twitter use that is not oriented to by users as problematic or subjected to efforts to repair. This goes hand-in-hand with describing exchange systems like Twitter as 'asynchronous', but it does also immediately render it as something quite distinct from face-to-face conversation.

Another observation made by Sacks et al is that "turn order is not fixed, but varies". For conversation the observation here relates to the fact that it is not continually ordered such that speakers have set and allotted turns, e.g. Speaker A / Speaker B in interrogation, but rather it is clear that speaker changes cannot be predicted in advance of the ongoing turn, and even then which speaker goes next is not always fully decided prior to the transition point between turns. This feature of face-to-face conversation is entirely concordant with the organisation of tweets in Twitter where which tweet falls next in the timeline is not predictable in advance. Thus you can get the following kind of pattern where a whole set of distinct tweeters all respond individually to an initial opening conversational gambit without any predictable relationship between them (see Figure 10.1).

Another point of interest here is that the fact that the turn order is not fixed results in a potentially indefinite number of people self-selecting to tweet in response to a prior tweet, the only constraint in operation being the size of the cohort of followers of the person who tweeted initially (with retweeting creating scope for endless extension of this cohort to other people's followers). The number of potential self-selecting next speakers in face-to-face conversation is tightly controlled both by the limits that exist on the number of people who can be co-present and in range to hear, and by a range of incumbent rights and obligations that exist as a feature of the relationship between the people who are co-present. This feature in particular is of moment for how Twitter can serve as a highly effective conduit for the transmission of rumours across non-related cohorts.

Something else that came out of the work of Sacks et al is the observation that "turn size is not fixed, but varies". Their comment with regard to this particular feature is that other turn-taking systems can be quite distinct from conversation in how turn size is handled. So in debates, for instance, one can see that the length of turns is quite tightly

¹<http://www.beevolve.com/twitter-statistics/>

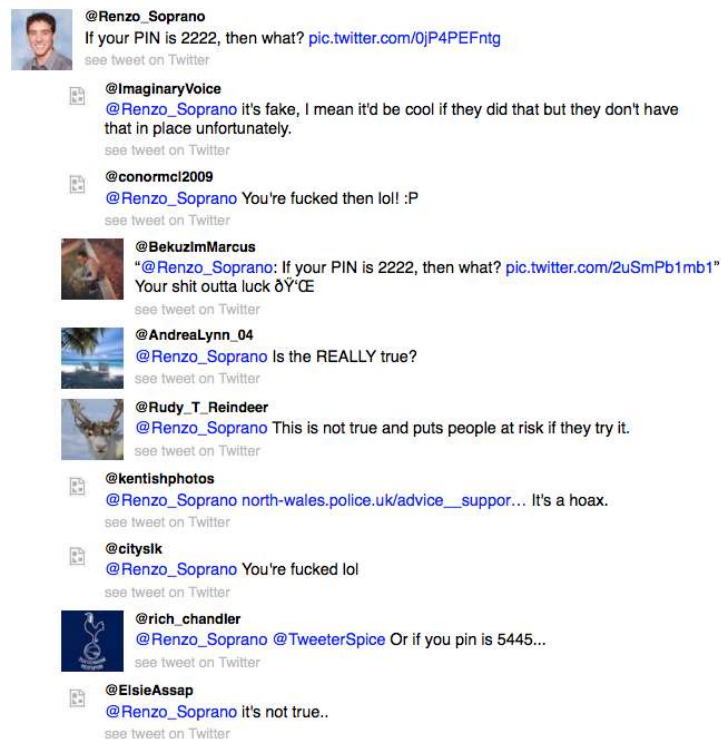


Figure 10.1: Tweeters all respond individually to an initial opening conversational gambit.

pre-specified. An important aspect of Twitter here is that, whilst the exact length of a turn may not be pre-specified, its maximum length is very tightly constrained at 140 characters, although strategies can be adopted that result in something akin to an extension of the turn. One such strategy is the linking of posts. Another is the completion of the turn over multiple posts, using the convention of three dots at the end of each post to indicate that there is more to come. However, it should be noted that in the context of the twitter-stream, as it is encountered by recipients (or, to be more accurate, 'followers'), these strategies still result in separate posts that look to all intents and purposes like separate turns. Thus these strategies, whilst managing to indicate that there is some connection between multiple posts from the same person, resonate more as topic reference markers (e.g. like 'but as I was saying before', 'with regard to...', 'coming back to...', etc.) in that there is a marker external to the actual content to be provided that makes evident to recipients the presence of a connection. Thus there is a sense in which turn-size is pre-specified in Twitter, or at least absolutely constrained.

Similarities between conversation and Twitter use do exist. For instance, Sacks et al make the observation that "turn allocation techniques are obviously used... a current speaker may select a next speaker (as when he addresses a question to another party); or parties may self-select in starting to talk". Despite its asynchronous character and the potential interleaving of a number of distinct sequences of tweets on Twitter it remains

the case that tweets are composed and arrive as distinct units within global Twitter feeds. With regard to any one particular topic there is a 'first speaker' in terms of there being an originator, there are subsequent parties who may be implicated as respondents within the original tweet, and there are parties who select themselves as respondents to a tweet in some way. Differences here particularly relate to other matters such as: 'co-placement', where responses to a specific tweet may not be sequentially directly adjacent to that tweet within a feed (because, in principal, all comers may respond to all tweets, so next up in a feed may be an entirely unrelated response to a different topic); and 'rights of response' in that any recipient of a tweet may respond to it or retweet it, whilst this is clearly not the case in face-to-face conversation, where just who gets to speak is a very tightly managed affair.

Something of moment across all sorts of organized human phenomena is the existence within them of practices to bring about repair. In the context of turn-taking mechanisms Sacks et al comment that "repair mechanisms exist for dealing with turn-taking errors and violations; e.g., if two parties find themselves talking at the same time, one [or both] of them will stop prematurely, thus repairing the trouble". In this respect they note that there is a variety of ways in which repair of troubles in the turn-taking system can be undertaken, including questions, apologies, repeats, stopping things dead before completion, and so on. They also make the observation that: "A major feature of a rational organization for behavior which accommodates real-world interests, and is not susceptible of external enforcement, is that it incorporates resources and procedures for repair into its fundamental organization." ([Sacks, 1978]: 720). An important implication of this is that, whilst it may differ in certain aspects of its realization, the organizational arrangements of tweet-exchange should also exhibit procedures for bringing about repair. The technical constitution of Twitter mitigates the prospect of physical overlap of turns occurring. However, as we shall be examining below there are a variety of ways in which tweeters are nonetheless held accountable for the content they produce and how it is motivated, which in turn can implicate the actual production of accounts or actions that amount to practices of repair (e.g. the withdrawal of a specific tweet, apologies, explanations, elaborations, and so on).

In the following discussion we will be looking at various more specific conversational phenomena. However, it should be noted that across all of them there are ways in which they are also ongoingly oriented to the kinds of turn-taking elements we have outlined above.

10.1.2 Topic

The organisation of conversation around topics, topical coherence, and shifts of topic is a central focus of the conversation analytic literature and understanding topic-based relationships is important for being able to track the flow of rumour-type phenomena across large bodies of tweets. Clearly responding to other people's tweets, commenting

upon embedded tweets being retweeted, and simple retweets all exhibit certain features of topical coherence, and Twitter itself also reflects this understanding in its grouping together of connected tweets in this way as 'conversations'. Grosser degrees of topical relation may also sometimes be encapsulated within the use of hashtags. Looking at some more specific examples we can see how topical coherence is both routinely handled in Twitter and a potential problem that has to be managed within an unfolding series of turns, giving rise to potential misunderstandings than can, given the nature of follower/followed relations in Twitter, result in the spread of potentially unverified content.

The following materials relate to the crash of flight AH5017 at a time when it was not yet evident that it had actually crashed and when speculation was rife. Flightradar24 is a well-known live flight tracking system that also provides social media commentary on both Twitter and Facebook. The conversation captured here begins with a response from flightradar24 to a query about a possible missing flight (see Figure 10.2).

ah5017-492234753060249601

The screenshot shows a Twitter thread starting with @flightradar24: "The only flight between Ouagadougou and Algiers today is #AH5017. We have no confirmation that this is the missing flight!"

Replies include:

- @jwadmnazir: "What do you mean missing flight? Another disaster? #MH370?"
- @flyhellas: "What aircraft is used on that route? A330/738/ATR?"
- @flightradar24: "Our schedules say that it could be a B736... Not confirmed!"
- @flyhellas: "Ok thankyou. I was worried it couldve been something bigger from their fleet. Thankyou. Lets pray the worst hasnt happened"
- @mike_juliet: "apparently its there leased MD83 operating the route today."
- @ortyzapple: "Sono stati persi i contatti con un altro aereo... Di Air Algerie a quanto sembra. Forse l'AH5017."
- @Miss_Twin_Peaks: "ma che brutto momento, eli? Davvero. Mi ricorda qualche anno fa. Ci fu un'estate del genere..."
- @cms66: "Is not a good year to flight."
- @Alextohechi: "The only flight between Ouagadougou and Algiers today is #AH5017. We have no confirmation that this is the missing flight!"
- @toktokalweert1: "It is still sceduled. #AH5017 pic.twitter.com/Y5MVIEDKX"
- @thecaptain707: "indeed strange, 5 hours or so late and still scheduled to arrive (orig arr time 0540 local)"
- @raphaelcockx: "Am I right in assuming you have no coverage in Burkina Faso?"
- @DanielSander95: "no more crashes please that's 3 in one week if confirmed missing. :-("
- @rjonesy: "map? No clue where this is"
- @mwyres: "Seems that it is #AH5017: af.reuters.com/article/algeri..."
- @helloimyour: "THAT IS HIM"

Figure 10.2: Conversation about a possible missing flight.

Now a lot of the tweets here are directed to @flightradar24's original tweet rather than to one another, though there are a couple of brief side conversations. @flightradar24 itself orients to almost all of the subsequent tweets as comments rather than turns to which it should respond, even though two of them, @raphaelcockx and @rjonesy, pose further questions. This demonstrates nicely something we have already noted above, namely

the right of followers to comment freely upon posts coming from the people they follow. Alongside of this we can see contributors such as @Alextobechi presuming a right to retweet the original tweet to their own followers. There are also a couple of tweeters who elaborate slightly upon the original tweet by pointing out that the proposed flight is still scheduled (e.g. @toktokalweer1ei and @thecaptain707). This then implicates a supporting elaboration from @toktokalweer1ei who provides a picture of the relevant flight board. @thecaptain707 now produces a turn that is both a comment upon the preceding tweets and an elaboration, by saying that skynews has just tweeted the same information. After this @mwyres responds more directly to the original tweet by offering something closer to confirmation that it is indeed the flight first proposed. Here they also provide additional support of their confirmation through a provided link to an article. Note also, however, how this is still provided with a limiting belief marker: it 'seems that it is AH5017'. @helloimyour, by contrast, presumes a right to give direct and unsubstantiated confirmation of the flight: 'THAT IS HIM'. Something in particular to notice here is that there are a range of returns possible to an originating tweet that go beyond face-to-face conversation because all of the recipients have an in principal right to respond in some kind of way, and without necessarily providing any account for the provision of a response. So, as a recipient, there is effectively an immediate presumption that it is accountably appropriate to respond such that no one sees the need to justify that. Similarly there is no presumed need to have been marked out as a next speaker in any way, or to justify your self-selection beyond the tweet you provide being somehow on topic. At the same time it is also presumed that having side-conversations occasioned by the original tweet is in no way problematic. However, it is also important to note here that the orientation to accountability for content of the response is potentially more constrained. It's okay for anyone to comment, but not in just any kind of way. Thus we see that some tweeters will use markers to limit their accountability for their response, even though others don't.

Important for the ongoing spread of tweets, something else to notice here is that there is a further presumed right to just hand on the information by retweeting without having to account for the passing it on in any way. There are limits here, which we shall return to below in our discussion of reportability: it is necessary that a retweet be seeably relevant to your followers. Retweets that do not have any status on the basis of being news or evidently interesting, novel, etc. might be more open to question. Having said this, a distinction between Twitter and some other social networking services is that followers do not necessarily know the interests of other members of the cohort of followers of a particular person, so there is a difficulty in calling to account in that it might have been retweeted for the benefit of some other follower, just not for you.

Following through still further on the issue of followers and the topical coherence of tweets, note here how @jawadmnazir's post quite early on in the stream, presumes a right to question the original tweet: "What do you mean missing flight? Another disaster?" This demonstrates how it is perfectly allowable within the system to pick up on the content of a prior post and call it to account in some way. The feature called to account here is @flightradar24's second sentence mention of the missing flight, echoing my prior

observation that it is possible that many followers saw this as a first post and sought to find an account for the odd second part. The feature of the system that gives rise to issues here is that when you follow someone you often get posts from them that are responding to other people's posts you are not privy to, so you effectively enter the conversation part way through and have to disambiguate what the features may mean. It is clear that for tweeters some understanding of topical coherence continues to be oriented to even though the structure of tweeter/follower relations may serve to breach that because the cohort of people being followed and the cohort of people following are not commensurate so not all parties are equally privy to all parts of a conversational stream. This leaves space for a presumption of meaning, retweeting, and reinterpretation of an original tweet without full cognizance of how a conversation may have unfolded, which in turn can give rise to certain kinds of misinformation. There is a mechanism for disambiguation because a post can be clicked on to expose the thread it is a part of, even if you do not follow all of the people in the stream whose posts are displayed. An ongoing question here is how many people do this? It also makes posts where the meaning is seemingly apparent but in fact bound up with a prior invisible exchange more open to being misunderstood. An analogy here is walking into a room part way through a conversation and thinking you know what is being spoken of after listening to the ongoing conversation for a while, taking a turn, and being pulled up short through the reiteration by others of what you were not privy to as a form of repair. e.g. "no, what we were actually talking about was...". However, once again because of the disconnect between 'follower' and 'followed' cohorts in Twitter, it is by no means certain that all parties will see subsequent posts to engage in this kind of repair.

A further temporal consequence of how Twitter is organised is that the time spans over which respondents may address themselves to a topic without loss of coherence are much greater in the case of Twitter than they are in face-to-face conversation. In ordinary conversation, as most speakers will readily recognise, failure to address oneself to a topic quickly enough means that another topic will be floored and addressing oneself to the original topic becomes much more difficult and accountable. Conversation analysis has looked closely at how 'change of topic markers' are handled in conversation. Part of this also relates to 'return to topic markers' such as 'but as I was saying...', 'but going back to what you were saying earlier about...', and so on. Thus, there are ways of managing topic preservation over more extended periods in spoken conversation.

The temporal organisation of Twitter provides for certain distinct but equally systematic ways of marking out topic relationships in order to manage coherence across more extended conversational threads. Twitter is, in fact, asynchronous in terms of response by nature, with a large number of unrelated tweets appearing moment by moment within the stream and with tweets addressed to the same topic being potentially widely spaced apart. Re-tweeting is one obvious way in which this is accomplished. Another more specific technique can be the direct mention of the previous speaker which has the dual effect of both indicating the presence of a topic relation to all witnessing parties and of ensuring that the person specifically addressed sees your tweet (see Figure 10.3).

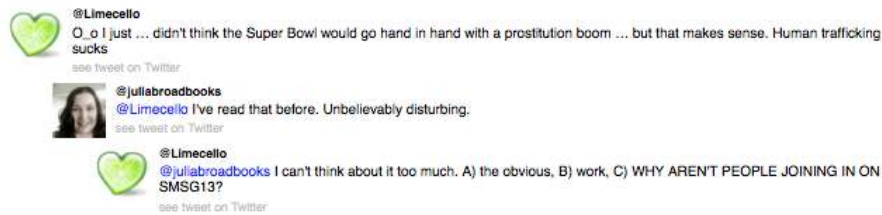


Figure 10.3: Example of a witness responding in a conversation.

Grosser degrees of topical relation may also sometimes be encapsulated within the use of hashtags. There are other indicators of 'on topic' / 'off topic' that can be seen to have a clear continuity with methods used in face-to-face conversation. For instance, note the use of 'btw' in Figure 10.4 and how the respondent handles both continuation of topic and the transition that has been proposed:

10.1.3 The organization of conversation as applied to tweets and the organization of tweets when seen as conversations

In order to make full use of the extant conversation analytic literature one of the longer-term activities necessary is to work through the principal organisational devices in play in conversation that conversation analysis has identified over the years, to explore how these devices might or might not be present in tweet-based phenomena in various ways, and to examine the extent to which they are organised in a similar fashion or otherwise. Such devices can be seen to include: adjacency pairs; change-of-state marking; correction-invitation devices; formulation; membership categorization devices; prefacing; premising; pre-sequences; receipt tokens; recipient design; repair procedures; sequential objects; speaker selection techniques; topic marking; and so on. Each of these areas of interest has a large body of literature already devoted to it. Some of the areas more evidently related to the concerns of the PHEME project have already been discussed above because of their foundational character (i.e. speaker selection and topic management). A number of others have potential relevance for the current annotation scheme and may therefore reward further investigation. Where relevant to the existing annotation scheme these are grouped under related headings, otherwise they can be seen to constitute ways in which the annotation scheme may subsequently be extended.

Factuality (Presentation/Claim)

Ambiguity: Whilst it is possible for a range of utterances to be taken as ambiguous with regard to their meaning, some research in conversation analysis also points to ways in which utterances can be ambiguous by design. This may have relevance with regard to how the factuality of claims is first presented in tweets with it being deliberately the case that people might take what is being claimed in several different ways. As an example, note [Sacks, 1995]'s comments on the use of the word 'you' where it could equally mean

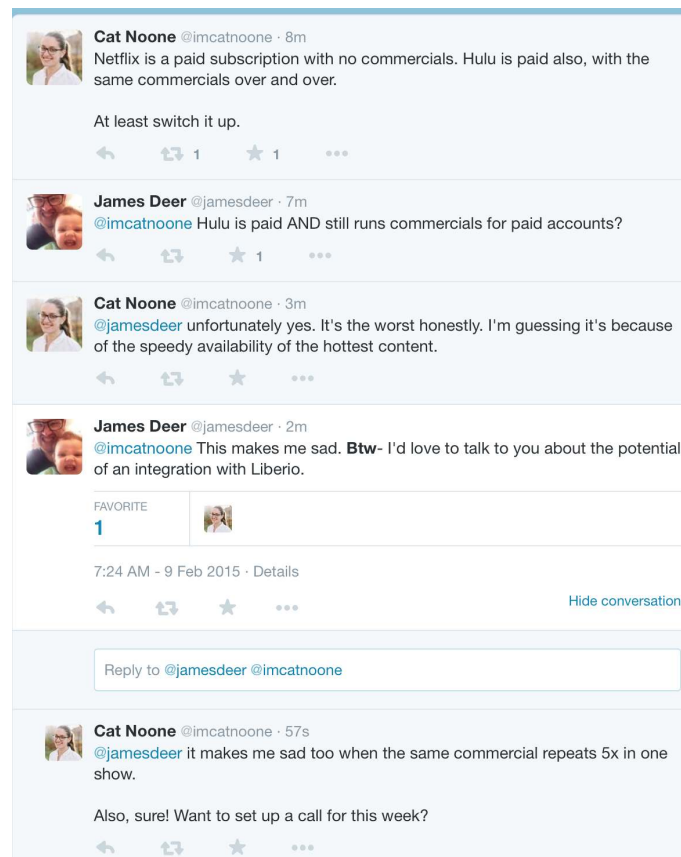


Figure 10.4: Example of a topic shift using 'btw' in a conversation.

'one' or it could mean 'they, e.g.': "If you're hotrodding you're bound to get caught". In this case the 'you' could be directed at the specific recipient, it could be directed at another body of people, or it could be a more general observation upon the outcomes of hotrodding. Even in the context of its production it was not clear and the recipient was obliged to settle on one understanding. With regard to rumours it should be noted that this kind of method stands as a) a resource for 'getting hold of the wrong end of the stick', but also b) a reasonable account for claiming 'they got hold of the wrong end of the stick' so that, if things turn out badly, it can be claimed that your utterance was taken the wrong way.

Framing, prefacing, premising markers and indicators of relative certainty: Another part of the conversation analytic literature relevant to the status of claims as presented refers to the ways in which different utterances may get framed or prefaced in order to inform recipients how to understand the speaker's orientation to what they are about to say. Many of these relate to matters of certainty, for instance 'I believe that...' (see [Coulter, 1979]), 'I think...', 'I thought...', 'I understand that...', 'it would seem that...', and so on. Utterances that may shape up to be rumours can include these kinds of pref-

acing words or remarks, e.g. 'Apparently the rioters are moving towards Birmingham Children's hospital'. It is important to note that these are not just about the certainty or otherwise of a speaker producing them, but also about providing for how the speaker might be called to account for what they say. As an example note in the conversation presented above regarding the disappearance of flight AH5017, how @flightradar24 responds to a question from @flyhellas regarding what kind of aircraft is used on the route by saying "Our schedules say that it could be a B736... Not confirmed!" Within this single tweet there are three methods adopted for offsetting potential accountability for this information and thus framing how recipients should understand its status. First of all it does not say 'it could be a B736', it says 'our schedules say...'. This immediately offsets any personal liability for the claim by framing the source as a set of schedules that are of unknown and potentially variable accuracy, rather than the respondent himself/herself. Secondly, the response is framed as 'it could be a B736', not, that it is a B736, further distancing the respondent from certainty about the claim and simultaneously indicating further the potential inaccuracy of the schedules. Finally the respondent places at the end of the tweet the words "Not confirmed!" to underscore how they wish the information to be taken. Thus we can see there are a range of to-hand linguistic methods for indicating the status of information that can be drawn upon and that would be visible in any subsequent retweets, limiting the scope for it to be advanced as a genuine claim and thereby acquire the status of a rumour. The very fact that these methods can be seen in how tweets are sometimes framed provides further insight as to how they might otherwise operate as rumours because it is the systematic absence of such framing devices that provides people with the scope to treat tweets as claims because the status of the claim is left open.

Evidentiality

In section 5 one of the features of the annotation scheme is evidentiality. This refers to the degree to which evidence is provided within a tweet to support the claims or propositions being made. The conversation analytic literature has also addressed itself to this kind of phenomenon and how speakers might go about producing utterances in such a way as to not be called to account for them being dubious in some sense. It explores the matter in a variety of ways:

Sacks [51], in a discussion regarding the distinction between claiming and demonstrating in conversation, looks in particular at the work that can be done by second stories. As we discuss again below with regard to motive power, a commonplace phenomenon is that when one person tells a story another person will follow it up with a similar story of some kind. If a first story is simply followed by 'I know just what you mean' or 'I agree' and nothing more this amounts to only being a claim that you are aligned with the speaker in some sense. Telling a second story that exhibits the same point from your own experience serves to actually demonstrate your concurrence. So there are methods for making clear that you are doing more than just claiming alignment that are oriented to as acceptable ways of doing that. In rumour production, then, second story production is one way in which speakers may demonstrate whether they attach credibility to the rumour in some sense.

In another discussion about the character of story production Sacks (op cit) makes an observation that is in some ways the counterpart of the observations above regarding ambiguity, which amounts to saying that 'brevity invites inference'. Sacks' point was that where speakers are concerned that a story may lead to the wrong kinds of inferences being made they will often elaborate the story quite significantly to ward off potentially awkward judgments (the specific case discussed related to potential assumptions regarding a man's sexual preferences). This is relevant to rumour production in a variety of ways. For instance non-specificity can deliberately encourage speculation. And one way of trying to encourage acceptance can be through the production of detail.

Another relevant discussion in conversation analysis relates to how people use certain kinds of stock phrases such as "everyone does that don't they" as a means of setting aside all further need for account. Proverbs can also be seen to be used in the same kind of way e.g. "better the devil you know". Once again, as a response to rumours, such phrases can be seen to be doing quite definite kinds of alignment work. In particular, once produced within a stream, they make it such that to contest matters now is not just contestation of the rumour but also contestation of a stock body of knowledge that has been applied, something that is much harder to do. Indeed, conversation analysts have catalogued quite a range of instances where something is made to be self-evident by its association with a particular thing that just anybody knows. In other words a routine way of promoting acceptance is to work something up as being just another case of what everybody knows.

Yet another body of work looks at the range of methods whereby the grounds of claims are made visible, where inference is supported or resisted according to need, and where the very need for evidence is set aside. Benson Hughes [4], for instance, explore how the work of variable analysis trades upon a range of ordinary competences and commonsense assumptions and how the recognisable adequacy of statistics as evidence trades upon these things. This can be seen to extend to ordinary everyday interactions where to produce a statistic is commonsensically seen to be providing a certain kind of claim regarding the credibility of what is said.

In the case of Twitter other kinds of evidential practices can be seen to be brought to bear. So to briefly indicate some of these practices, returning once again to the discussion of flight AH5017 above, whilst we see some tweets carefully putting bounds around the kinds of claim that might be being made, there are others which directly seek to present supporting evidence for their position instead. Thus we see @toktokalweerlei offering up potential counter-evidence with the observation that the flight is still scheduled and an accompanying link to a photograph of a flightboard displaying the flight information. @thecaptain707 responds to this with an observation that whilst it may still be scheduled to arrive it is now "5 hours or so late". He follows this up some 45 seconds later with the comment that Sky News has just tweeted the same information, which, as with the photograph, amounts to an appeal to sources beyond the tweet itself for the grounds of the claims being made. This is reinforced about 20 minutes later by @mwyres posting a link to an external Reuter's article supporting the proposition that it is flight AH5017. The important thing to note here is that, unlike the situation with face-to-face conversation

where evidence for claims is something that has to be methodically handled through the further production of talk, Twitter offers up to tweeters the capacity to link their tweets to other external resources as a methodical way of handling the production of evidence for their claims. In other words one can find a series of referential practices in the use of Twitter that are not produced in such explicit ways in spoken conversation. Whilst this in no way guarantees the validity of the referred to resources, it does provide tweeters with a means of making claims seemingly more compelling, which is another potential building block for the power of Twitter to promote the spread of rumours and speculation.

Another systematic feature of Twitter that makes it distinct in some ways from face-to-face conversation and that is also of pertinence for considerations of how people may orient to rumours content, is the way in which location may feature with regard to the possible credibility of claims. We discuss below under the topic of 'lying' the ways in which it is incumbent upon conversationalists to produce certain elements within their accounts that can testify to their witness status for those accounts. Now conversations reporting something that happened are most often occurring after the event and at some remove from it physically. However, Twitter provides for the possibility of producing accounts of events whilst they are actually in the course of taking place. A feature of the evidential status of such tweets is therefore the extent to which the identifiable location of the tweeter is commensurate with the claims being made. Twitter does allow for location to be made known which obviously makes this more explicit, though it is not a feature that is by any means always made available by tweeters themselves. People may also be evidently tweeting from within their homes or other specific locations by virtue of the content such as express descriptions of activities routinely assumed to take place in certain places such as sitting down to a Sunday dinner or watching TV, having a drink with friends, etc. Locations may be specifically mentioned within tweets as well. Photographs may also be understandable as indicators of location.

Warrants: Related but in some ways distinct to the preceding discussion but nonetheless still bound up with matters of evidentiality is the matter of warrants or rights to be able to claim certain things in certain kinds of ways such that what is said is taken for granted to be true. Discussions here lead to something we shall be discussing in greater detail below which is that just how people are categorised in talk already sets up a bunch of assumptions regarding what might be reasonably claimed about their actions (and thus never called to account in any way). Sacks [51], for instance, discusses a report of an incident where part of the report is that the sister calls the police. He points out that within the report and the response to the report the nature of the sister (is she elderly? is she prone to hysterics?) is never put into question. Part of the nature of categorisation of people is that it provides for warrantable action, e.g. as we shall be seeing below, Hell's Angels rape young girls and Hotrodders like to drive fast cars. Sacks makes the strong claim here that "a task of socialization is to produce somebody who so behaves that those categories are enough to know something about him". However, he also points out that these kinds of assumptions are overturn-able as assumptions by other rights of precedence, e.g. witness status or local knowledge. With regard to rumour a case in point

here is the following extract from the London riots tweets and the rumour that rioters were attacking a children's hospital and that police were massing to protect it: May I remind clueless/hysterical birminghamriots commentators that Children's Hospital sits face-face with city's central police station.

With regard to all of the matters we have discussed above an important element to hang on to here is that people methodically build into their utterances from the word go ways in which they might or might not be held accountable for the production of those utterances. So at least one part of the work of unpicking matters of evidentiality and plausibility and acceptability and veracity in sequences of tweets is to look at how people are systematically managing their accountability in the production of those tweets.

Plausibility

When it comes to matters of both plausibility and acceptability there are once again a variety of conversation analytic treatments of such matters that may assist with providing possible insights and points of comparison.

Lying & Veracity or Truth: There are a number of discussions in the literature regarding lying and truth. The central outcome of analysis here is that there are routine grounds upon which the prospective character of something as either a 'lie' or a 'truthful' account may be established. Sacks [51], when discussing the production of competence in the telling of a story, looks at a report of a car wreck to observe how the tellers make it evident that they have the competence to be reliable witnesses of car-wrecks such as 'we were stopped there for 25 minutes' and 'the car was smashed into such a small space'. Sacks' point is that people have a sense of what's usual for the report in play. Thus, stepping outside of that can prompt the questioning of its truthfulness e.g. 'we were stopped there for just a second'.

Subjectivity & Objectivity: Another related matter here is how people work with, on the one hand, what just anyone knows of the world, and on the other with what only certain people in certain positions might know of the world. Much of the conversation analytic literature points out that lots of tellings trade upon what just anybody knows of the world such that the claims made might, as a routine supposition, be seen to have an objective character until such a time as it might be there are grounds for thinking otherwise. Tightly bound up with this is Sacks' discussion of 'Doing Being Ordinary' [50]. His observation is that, for any activity, there is a presumed ordinariness about what is going on. People make commonsense assumptions about what the ordinary business of any state of affairs might be and only pause to remark upon things that fall outside of that. The implication of this is that there are ordinary ways of having riots, the same as anything else. Ordinary expectations about riots would include things like places being set on fire, guns being shot, policemen beating people, such that images of such things would not necessarily invite inspection. Thus the scope for spread of a rumour and the chances of it being called out trades upon there being background expectations in play such that the things being proposed fall within the scope of being the ordinary business of stuff like that. And it is exactly when, for instance, an image falls outside of such background expectations that it

is subject to remark, open to inspection and potentially rendered in need of an account.

Acceptability

Trust: A further elaboration regarding the points we have made so far is that the accountability of people also typically comes with notions of trust and rights and responsibilities built in. Of particular moment here is the matter of reporting to known others versus overhearing and getting stories from unknown others. So, for instance, saying to someone 'Well Sammy told me such and such', where the other party also knows who Sammy is, provides systematically for: the accountability of the speaker to Sammy and to the person they're talking to; for the accountability of Sammy for just what he said to these parties; and for the receiving party as well as to just what they might then be moved to report. However, overhearing falls outside of these routine arrangements of accountability and trust. So, an overhearing party can just report the such-and-such that was overheard without the need to make those accountabilities visible. They are only required to provide provenance if explicitly called to account. With regard to rumours on Twitter note that, for many Twitter retweets, people are passing on tales from unknown others so they already stand outside the routine arrangements of trust and accountability.

Membership Categorisation Devices (MCDs)

This refers to a strong orientation people display towards hearing certain things that might be heard as going together as indeed going together. The phenomenon was first described by Sacks [47] as a feature of the analysis of stories told by children. He pointed to the strong tendency of native English speakers to hear the utterance "The baby cried, the mommy picked it up" in such a way as to understand that it is the mother of the baby who picks it up, even though this is not actually specified. He elaborated upon a range of membership categorisation devices together with a set of tying rules (not actually 'rules' in fact but rather maxims) that provide for how people hear things as going together. In another discussion of MCDs [Sacks, 1979] discussed how different categorisations of exactly the same people might be used to do moral work. Thus teenagers might refer to one another as 'Hotrodders' (with certain 'cool' connotations), whilst adults might refer to them as 'kids in cars'. This then provides for taking quite different positions regarding the matter of driving fast. Slightly later work on MCDs has often focused upon examples closer to Sacks' Hotrodders where there is a deliberate use of 'morally contrastive categories'. [Lee, 1984], for instance, in a paper entitled *Innocent Victims and Evil-Doers*, discussed in detail the newspaper headline "Girl Guide Aged 14 Raped at Hell's Angels Convention". Here the categorisations deliberately provide for seeing the parties involved in highly distinct ways. In the context of rumours it is likely that the latter kind of MCDs, drawing upon morally contrastive categories, will prove more fruitful for inspecting how both the crafting and spread takes place. Points of particular interest include how MCDs can provide for naturally presumptive work such that certain kinds of tweets may go unchallenged, e.g., for the London Riots data 'Rioters set Miss Selfridges on Fire' is altogether less remarkable and open to inspection than something like 'Grandmother sets Miss Selfridges on Fire' would be; and amongst certain communities 'Police

beat a 16-year-old girl' is potentially more credible than something like 'Councillors beat a 16-year-old girl' might be.

Sequential Ordering

Another aspect of conversation analysis looks specifically at how the positioning of utterances in relation to one another can serve to inform the ways in which they are taken to be meaningful. Sacks [48] engages in an analysis of the telling of a dirty joke in order to illustrate this. He works through how the assembly of a set of potentially un-related utterances into a specific sequence can invite a certain reading where to get the joke is to see that reading and find it funny. This may be relevant for work on rumours in terms of how the crafting of specific messages may be taken to be implicative and also in terms of how to assess different kinds of response, e.g. (from the London Riots data): 'Apparently McDonalds stormed in tottenham. Rioters proceeded to take over and cook some burger 'n fries. Ya can tell it's school holidays'.

[Sacks et al., 1974] expressly examine how people orient to turns as 'turns-in-a-series' in conversation:

"Turns display gross organizational features that reflect their occurrence in a series. They regularly have a three-part structure: one which addresses the relation of a turn to a prior, one involved with what is occupying the turn, and one which addresses the relation of the turn to a succeeding one. These parts regularly occur in that order, an obvious rational ordering for an organization that latches a turn to the turns on either side of it." (ibid).

This, in turn, has important implications for how analysis to turn-taking may proceed as an analytic enterprise:

"... while understandings of other turns' talk are displayed to co-participants, they are available as well to professional analysts, who are thereby afforded a proof criterion (and a search procedure) for the analysis of what a turn's talk is occupied with. Since it is the parties' understandings of prior turns' talk that is relevant to their construction of next turns, it is their understandings that are wanted for analysis. The display of those understandings in the talk of subsequent turns afford both a resource for the analysis of prior turns and a proof procedure for professional analyses of prior turns - resources intrinsic to the data themselves." (op cit: 725).

In other words, the local analysis of a prior turn that is made visible in a subsequent turn is itself a resource for our own analysis. It tells us what the participants to a course of interaction understand to have been done themselves at every step of the way. This underpins the position we have adopted in the annotation scheme whereby we argue that what counts as a rumour is what is manifestly taken to be a rumour and handled that way in the turn that follows what is seen to be the source of the rumour in the first place. Thus there is no point in looking to any one turn and seeing it as amounting to a rumour in any free-standing way. It is the turns that follow that will be seen to matter.

An additional feature of the sequential placement of tweets in Twitter relates to how

Twitter is an ongoing stream of tweets that may be posted and encountered at any time of day. This being the case, the timeline on Twitter is a powerful temporal ordering device that quickly becomes a resource for reasoning about posts for those who use it regularly. It also provides for the temporal organisation of sequences of interrelated posts, despite the accumulation of other materials between them. In this sense the timeline and the visible temporal placement of posts within that, complete with time-stamping and prospective immediate visibility to one's cohort of followers, makes time of day an important constraint with regard to what might or might not be accountably posted. Some posts, for instance breaking news, can be circulated at a wide range of times. Indeed, here it would be more accountable to wait to post because the whole thing with breaking news is that it's posted as soon as you hear of it, not wait until some convenient later moment. And there are a range of newsworthy topics where immediate telling is expected and where those who know you might reasonably ask 'why the delay?' if you do choose to wait to circulate it. This is much broader than Twitter but is also a relevant feature of the organisation of people's Tweets. However, there are other things such as prospective reading on the way to work, tales of pub exploits or gigs, recountings of lunches, or even the posting of recipes, that have certain temporal expectations attached such that, for instance, posting roast dinner recipes at 5 in the morning might reasonably prompt some of one's followers to ask 'why that now?'

Another important aspect of sequencing in conversation is the way in which certain elements are tightly bound together in pairs, termed in the literature as 'adjacency pairs' (see [Sacks, 1995]). So, returning once again to the materials above regarding flight AH5017, note that when @flyhellas posts "@flightradar24 What aircraft is used on that route? A330/738/ATR ?" not only does this indicate a presumption by this follower that @flightradar24 is possessed of certain kinds of air traffic competence, it also makes @flightradar24 accountable (and thus responsible) for the production of a reply. These posts are not, within Twitter, obliged to be literally adjacent in either the stream or the conversation, but questions like this do still carry the power of implicating some kind of response. A critical thing here is the way in which the tweeter makes use of a mention by directly addressing @flightradar24. It can be seen elsewhere in the stream that there are questions raised that receive no response. @raphaelcock, for instance, produces a post that apparently implicates a response from @flightradar24 that, at least to the end of the excerpt about 1 hour and 40 minutes later, has received no response (unless via private DM). This is interesting because of both the lack of response and the ordinary assumption as a competent user of Twitter that a response to a question can take some time to appear without there being any necessary accountability for that. The responsibility is therefore one of producing an answer, but no longer adjacently or even, necessarily, any time soon (though this begs the question of just how long would people begin to treat as too long? Of course, a mention of @raphaelcockx and some part of their question would be enough to preserve the accountable relation by making the tie apparent. This is a distinction from verbal communication in that the resources to re-establish relevancy of response are harder to come by. One can say, 'oh, about that thing you were asking ...' etc., later in a

conversation and produce the re-binding effect. You can even do it much later by saying 'about that question you asked me earlier/ this morning/last night etc.' and it would still work as long as some reminder of the question itself was also produced. Thus it would appear to be the case that mentions have a particular status as a resource in this kind of communication for both rendering others responsible and/or accountable for a response and for demonstrating the binding of a response to some previous post in the timeline.

Reportability & Motive Power

One potentially important aspect of conversation analytic investigations regarding how people manage topics in conversations is the matter of how topics can get presented in the first place and, in particular, the notion of 'first topic status' [51] and how certain topics may count as 'news'. Conversation analysis points to how certain topics that are somehow remarkable or worthy of note provide people with the special licence for comment and retelling without the topic having been already implied by something else. This raises the question as to what counts as mundane or remarkable in what kinds of situations with regard to different kinds of social media – especially where there is a clear licence to report the otherwise mundane in certain ways.

What may be of especial concern here is what Sacks and certain other conversation analysts term 'motive power'. This rides on the observation that for most kinds of topic-raising some kind of account is routinely required. The account may often be self-evident because of other surrounding events or preceding utterances. However, some kinds of accounts are generative in their own right. Motive power refers to the extent to which stories and accounts are open to transmission to other people. One of the matters that impacts upon motive power is what Sacks [48] terms 'investment'. Investment refers to the degree to which relationships with people carry with them certain rights and obligations. So complete strangers show very little investment in one another, work colleagues may exhibit an interest in your health or where you are going for your next holiday but are unlikely to ask detailed questions about your love life, whilst daughters of a certain age are expected to report most things to their mothers but not necessarily vice versa, and husbands and wives are expected to tell one another pretty well everything. The upshot of this is that the number of people to whom you can report having met someone you haven't seen for a while on the way to the shops is very limited, whilst having seen a building on fire is much more widely reportable, and there are certain people who must be told certain things or trouble will surely follow, e.g., telling your mother you're getting married.

Another feature of motive power is what Sacks called 'entitlement to experience'. His observation here was that stories and jokes etc have high motive power according to the extent to which they convey experience. This is especially about the conveyance of experience that is out of the way and not otherwise available to you because you can figure the sheer remarkability of it is a thing that will make it self-evidently appropriate to report it. You are entitled to share it and other people are entitled to hear about it, which is not, of course, the case with just any experience you may wish to relate. A secondary phenomenon that relates to this that is also of interest is the commonplace expectation

that a telling of a story will prompt the telling of a second story in return by the recipients. This second story is routinely understood to need to be a telling of something similar that either happened to you or that you once heard tell of. It is also a primary way in which conversationalists demonstrate alignment with one another in their views upon different topics.

The upshot of all this is that matters that can be easily reported across different cohorts of people are possessed of certain characteristics. Typically they are matters that are remarkable or at least sufficiently out of the ordinary or representative of some noteworthy change that they merit report. In this respect they can be seen to overlap largely with matters we would be inclined to call news. An important additional characteristic of matters that are recognisable as news is the fact that they have what conversation analysts call a 'first-topic status': that is, they can be used independently of other ongoing topics of conversation to open new conversational threads and delivered apropos of nothing. They do not usually need to be prefaced, premised, or implicated by other matters in any way but can be said straight out of the blue without anyone calling you to account for it². This is clearly important for the transmission of rumours which are also often possessed of news-like characteristics and which are therefore open to: ready articulation as topics in their own right without special work being undertaken to make space for their production; articulation to members of other cohorts who you might not otherwise know.

This being the case we can observe that matters of reportability, motive power, re-tellings and alignment are all of potential significance for the spread of rumours. Some things that might be rumoured are clearly unlikely to carry relevance for anyone outside of highly constrained cohort of people (thus the potential distinction offered above regarding 'gossip'). However, other rumours convey matters that are tellable to a much broader set of people.

10.1.4 The intersubjective constitution of tweeting as a phenomenon

Something already alluded to above is the important to both conversation analytic and ethnomethodological approaches of the way in which any body of social accomplishments is an intersubjectively constituted set of accomplishments. These are reflexively organised around the specific understandings of the parties to those accomplishments of just what it is they are in the business of accomplishing. Furthermore, any specific feature is indexical of those mutual understandings in play. This may seem rather densely expressed but what falls out of it is that, to understand what is being done with any one particular utterance (or other kind of action) by one party, you need only look to the immediately subsequent utterance (or action) by the next interactant to see what kind of an action the preceding utterance has been understood to be. And, where misunderstandings occur (which, of

²Though it should be noted that they do still need to honour the sequential tying rules of conversation by being marked out as a new topic, e.g. by saying "Oh, by the way, have you heard...", "That reminds me, did you know that...", and so on.

course, they do) one need only look on to the utterance after that to see the original party engaging in some kind of repair. Thus interactants involved in a course of action routinely make available to others, who have the competence to see it, just exactly what is going on. Both CA and ethnomethodology trade in bringing this local reasoning into view. Thus they are occasionally called 'postanalytic' enterprises because they primarily work to make more explicit analysis that has already taken place on the part of those who originally produced the phenomena they are examining. The implication of this (and a significant challenge) is that annotation schemes truly aligned with conversation analytic and ethnomethodological approaches would seek not to tag text with externally derived analytic categories but would rather seek to identify the ways in which any specific tweet (or comparable phenomenon) has been analysed by members themselves in directly subsequent tweets in order to tag it appropriately. In particular, a focus upon clusters of 2 or 3 interrelated tweets is likely to be fruitful: initial tweet, responding tweet, subsequent tweet by originator (if there is one). This is a feature already being exploited by the annotation scheme we have devised.

10.1.5 Following and followers

Something else that falls out of the preceding observations is that it is important to understand properly the subtle mechanics of following/follower relations on Twitter so that just how their respective activities are aligned with one another and implicative for one another can be properly explicated. In particular, drawing upon observations first made in section 6.1.1, we need to note and be able to properly handle the fact that there are, variously: i) equal part conversations between parties who are mutually following one another, but also ii) audiences of interchange, who follow but are not followed, but who can nonetheless both comment upon witnessed exchanges and re-circulate those exchanges amongst their own community of followers, and, additionally, iii) subtle understandings in play of just who is following you, who your actions might be visible to, and how you might or might not be accountable to those parties in various ways. As a first step towards being able to capture this order of detail author types are already being displayed within the annotation scheme in order to inform reasoning about the status of tweets provided.

10.1.6 Tweeting as a mode of communication

As one works through the preceding body of materials something that becomes important to recognise is that tweeting is its own form of communication. It is not really conversation as in the sense of the classic forms of dyadic conversation that are the primary focus of conversation analysis. Nor is it good policy to simply assume that tweeting is just a specialised variant of traditional conversation in some way. Rather tweeting (or microblogging to use a slightly more formal term) should, in the first instance, be examined as a phenomenon in its own right with its own orderly characteristics that may or may

not prove to be tightly aligned with other kinds of communicative practices. Thus the safest approach is to take the corpus of findings coming out of CA as a starting point for reflection because conversation is a relatively well-described phenomenon and tweeting is not, rather than simply assuming that tweeting will operate in much the same way.

In this regard, there is a need to examine how identified 'conversational' phenomena within tweeting practices work as locally accountable features of a moral order. That is, as with any body of practice there are right and wrong ways of going about doing things and not just anything goes. Thus what happens within tweets may sometimes get explicitly called to account by other tweeters. Tweeters may themselves offer up 'accounts' for why they are proceeding in a certain fashion. Furthermore, one may test the orderly constitution of tweeting practice by deliberately exploring how it might be otherwise and what the consequences of doing things differently would be. All of these may serve to expose the socially mandated character of tweeting as a body of practice and how tweeters themselves manage it as an orderly set of affairs.

10.1.7 Looking at microblogging as its own job of work with its own grammars of action

In one of his most formative and programmatic papers called 'Notes on Methodology', Sacks makes the following methodological observations about how he first came to be working with talk and conversation:

"When I started to do research in sociology I figured that sociology could not be an actual science unless it was able to handle the details of actual events, handle them formally, and in the first instance be informative about them in the direct ways in which primitive sciences tend to be informative - that is, that anyone else can go and see whether what was said is so. And that is a tremendous control on seeing whether one is learning anything.

"So the question was, could there be some way that sociology could hope to deal with the details of actual events, formally and informatively? One might figure that it had already been shown that it was perfectly possible given the vast literature, or alternatively that it was obviously impossible given the literature. For a variety of reasons I figured that it had not been shown either way, and I wanted to locate some set of materials that would permit a test; materials that would have the virtue of permitting us to see whether it was possible, and if so, whether it was interesting. The results might be positive or negative.

"I started to work with tape-recorded conversations. Such materials had a single virtue, that I could replay them. I could transcribe them somewhat and study them extendedly - however long it might take. The tape-recorded materials constituted a "good enough" record of what happened. Other things, to be sure, happened, but at least what was on the tape had happened. It was not from any large interest in language or from some theoretical formulation of what should be studied that started with tape-recorded

conversations, but simply because I could get my hands on it and I could study it again and again, and also, consequentially, because others could look at what I had studied and make of it what they could, if, for example, they wanted to be able to disagree with me...

“Thus it is not any particular conversation, as an object, that we are primarily interested in. Our aim is to get into a position to transform, in an almost literal, physical sense, our view of “what happened,” from a matter of particular interaction done by particular people, to a matter of interactions as products of a machinery. We are trying to find the machinery. In order to do so we have to get access to its products. At this point, it is conversation that provides us such access. . . .”, Sacks [49]: 26-7

So something else to take note of here is that, just as Sacks was able to explore the production of certain facets of social interaction in a replicable and inspectable way by using tape-recorded conversations, so we have available to us within PHEME an equally replicable and inspectable body of recordings in the shape of a stream of tweets coming out of Twitter. There are some limitations here in that we do not have available to us the specific individual situation in which people composed and received those tweets.

However, Sacks’ original tape recordings were similarly constrained in that a good deal of ‘what was going on’ was absent from the recordings as far as the specific individuals being recorded were concerned. So what we do have in the corpus of tweets is a body of live-when-recorded socially produced phenomena that are open to being examined for how they work as – just as Sacks put it – ‘products of a machinery’. It is also worth noting here that, just as Sacks was concentrating on the orderly products of verbal interaction (often, it turns out, through the mediating technology of the telephone), so it is important that we focus upon the organisation of tweets in a Twitter stream as orderly products of an online interaction and focus on their own organisational properties as ‘just that kind of thing’, together with how those properties are made manifest and accountable within the way they are produced.

Elaborating a little on the preceding points, in that case, when you tweet you don’t typically say you’re just going to chat with someone, talk with someone, speak to someone, etc. How people articulate having conversations with one another and how they articulate tweeting, or even just looking on Twitter, are quite different. Throughout his work Coulter [8] makes much use of the notion of what he terms ‘sequential grammars of action’. This idea in outline actually originated with Wittgenstein. Wittgenstein [67] emphasises in his *Philosophical Investigations* that a grammar is in no way an explanation of action. It sets aside questions regarding why people do what they do. Instead it allows for us to see what resources they have available to them in particular situations and how they use them: “Grammar does not tell us how language may be constructed in order to fulfil its purpose, in order to have such-and-such an effect on human beings. It only describes and in no way explains the use of signs.” (Wittgenstein [67], PI: 496, p 138e).

In that they address questions of ‘what’ and ‘how’, the grammars of action in play when people are using Twitter are important. People ‘tweet’, they ‘retweet’, they ‘look at Twitter’, they ‘catch up on Twitter’, and so on. A job that therefore needs to be done is to

pull out of both Twitter and other sources of reference these different grammatical articulations of what people understand themselves to be doing when they are using Twitter and to lay these out as the structure of a body of practice. This should then provide for specifying how the different aspects of that body of practice that are articulated through these grammars are actually accomplished, what their orderly features of production look like, how those are in turn implicative for further bodies of practice and action and what those in turn look like. In other words it provides us with an understanding of the sequential organisation of people's practices that is grounded quite specifically in the use of Twitter, rather than taking those sequences to be a species of conversation that is primarily intelligible through reference to situated talk. There is a sense in which this work is preliminary to other work towards the development of a framework. However, pragmatically it makes sense to work with this perspective alongside the insights from conversation analysis because these are already to hand, working under the proviso that there will be a process of refinement and revision over time as our understanding of Twitter use in its own right develops.

10.1.8 The asynchronous character of microblog exchange

One of the more distinctive features of Twitter and microblog exchange in general is its asynchronous character. This forms one of the most important differences between Twitter and face-to-face conversation and some of the consequences of this have already been indicated, for instance, the absence of necessarily adjacent relations between related actions and the interleaving of different topics. As this constitutes such a significant difference it indicates a need to also examine closely how Twitter users systematically provide for its coherence across asynchronous interaction within the production of their own actions. Indeed, we have seen through our testing of the annotation scheme how this coherence is made observable and hence recoverable through the use of Twitter messaging conventions, in particular, retweet, reply and mention.

A crucial difference between the model of conversation that [Sacks et al., 1974] came up with and the situation regarding Twitter, that arises from its asynchronous character and that speaks to the nature of the phenomenon being addressed itself, is the way conversation is organised to provide for the minimization of gap and overlap. Conversation unfolds in co-present and linearly conjoint interaction such that gaps and overlaps are disruptive to the effective realisation of conversational talk. Tweets, by contrast, are textual productions that are, by virtue of the technical apparatus that enables them to be produced, both contiguous and without overlap. Thus this is not a problem to which the construction of tweets needs to be addressed. What one does encounter in Twitter, in particular in the context of what might be assembled by Twitter itself as a conversation, are phenomena such as: 'the complete absence of a turn', that is to say a turn by a certain party may be projectible but not forthcoming (we saw examples of questions without answers in the AH5017 conversation above); 'the conjoint production of largely unrelated turns', that is to say, two (or more) followers may set out to respond to an immediate prior

simultaneously in distinct ways, with the construction of Twitter resulting in their posts being assembled in the timeline consecutively even though they have no relation at all to each other but only to the prior turn (this, too, is evident in the AH5017 materials)³. It thus falls to the recipients of the conversation to disambiguate the relationship of the various turns to one another without the availability of their sequential production standing as a resource for such disambiguation (as it would in conversation), outside of the gross fact that certain turns can be seen to precede others and that a turn will, by necessity, be addressed to some other turn that precedes it rather than one that comes after it in the timeline. The importance of this distinction between twitter-based conversation and actual co-present conversation needs to be stressed. When [Sacks et al., 1974]: 715 delve into part of the issue of why talking at once might be a problem and how the turn-taking system provides an economical method for handling that problem, they discuss in particular how the model, by providing for the analysability of a turn of talk over the course of its production, might be impaired if turns were allowed to overlap, making projection of completion difficult to accomplish⁴. Twitter has effectively obviated the need for the turn-taking system in operation to handle this kind of problem by making it technically impossible for there to be overlapping turns.

However, in that Twitter use has moved beyond the original conceptualization by its originators of something that was largely designed to effect information exchange, and towards something that is oriented to as a device for specific user-to-user interaction over extended turns (as is recognised in the way Twitter now clusters related posts as conversations), the preceding observations also present a unique challenge to Twitter as it is currently constructed and the extent to which it can really be seen to operate as a conversation because it does not provide in the same way for systematically projectible conversational turns.

A further matter that is crucially bound up with the production of conversation in co-present interaction is the ongoing analysability of an utterance in the course of its production for its projectible point of completion⁵ and for the kind of work that is being

³Whilst it is organizationally different from this in a number of respects, [Sacks et al., 1974]: 712 do observe that the model they are proposing is foundationally geared to turn-taking in dyadic conversation with just two parties and that the addition of other parties can ramify. One of the ramifications they point to is that, when there are four or more parties, the talk can split up into more than one concurrent conversation with divergent talk happening at the same moment in time.

⁴In relation to, and in support of this they note that one kind of overlap is routinely acceptable in conversational exchange: “With regard to the ‘begin with a beginning’ constraint and its consequences, a familiar class of constructions is of particular interest. Appositional beginnings, e.g. well, but, and, so etc., are extraordinarily common, and do satisfy the constraints of a beginning. But they do that without revealing much about the constructional features of the sentence thus begun, i.e. without requiring that the speaker have a plan in hand as a condition for starting. Furthermore, their overlap will not impair the constructional development or the analysability of the sentence they begin. Appositionals, then, are turn-entry devices or pre-starts, as tag questions are exit devices or post-completers.” ([Sacks et al., 1974]: 715)

⁵Thus [Sacks et al., 1974]: 709 also note how variable turn-length is itself partly constituted by the nature of sentential constructions, which may themselves be extended through the inclusion of sub-clauses etc., and, in addition, comment that: “Sentential constructions are capable of being analysed in the course

done, such that a next speaker can be identified, can know when it is appropriate to speak, and can know what kind of a thing their own turn might need to accomplish⁶. Indeed, Sacks et al go so far as to suggest that the organisation of the turn-taking system may even key on all turns of talk having “points of possible unit completion . . . which are projectible before their occurrence” ([Sacks et al., 1974]: 716). They justify the proposed importance of people’s orientation to this, on the basis of the empirical materials they have accumulated, by saying that:

“Examination of where . . . ‘next turn starts’ occur in current turns shows them to occur at ‘possible completion points’. These turn out to be ‘possible completion points’ of sentences, clauses, phrases, and one-word constructions, and multiples thereof”. ([Sacks et al., 1974]: 717).

Clearly, recipients of tweets are able to engage in a post-analysis of the whole turn at their leisure, even to the point of re-examining it multiple times, before they complete their reasoning about such matters, and without the pressure of needing to step straight in when up-and-coming completion of an utterance is recognised. This lack of in situ pressure to analyse and respond also renders tweeting distinct from certain other kinds of text exchange such as live chatting and it provides for some of Twitter’s most unique organisational characteristics, including the scope for topically-bound conversations to unfold over very extended periods of time.

10.2 The organisation of rumour as a feature of microblog exchange

In this section of the chapter we are going to move on to specifically examining examples of already annotated tweet-exchanges where certain rumour-relevant characteristics have been identified. Moving beyond the basic annotations we are going to discuss specifically the organisational characteristics of the tweets in terms of the social science-based framework of analysis we have outlined above. This analysis will seek to move us beyond the primarily linguistic-based articulations of the rumour types based upon the existing annotations and towards unpacking what some of these conversations might amount to as social phenomena of various kinds.

As described elsewhere in this the deliverable, during the annotation process rumours can be browsed in two ways: by accuracy (true, false, unverified) and by acceptability

of their production by a party/hearer able to use such analyses to project their possible directions and completion loci.”

⁶Here [Sacks et al., 1974]: 710 point to the fact that, whilst ‘what parties say is not specified in advance’, certain kinds of turns do pre-figure what may thus be done with a subsequent turn, even if its exact content is not pre-specified. They additionally note that this feature can have an impact upon speaker selection in that certain types of turns pre-figure who the next speaker might be and what it is incumbent upon them to do.


(speculation, controversy, agreement). We will look briefly here at each of these potential understandings of rumours and their outcomes from a conversation analytic perspective to elaborate further some of the methodical practices and kinds of reasoning that are visible. Within specific examples it should be noted that the following additional features are displayed: a) how the posts have been annotated; and b) information (where available) about the 'actor types' for the individual tweeters, giving confidence levels ranging from 1 (lowest) to 7 (highest), whether they have been verified or not, and whether they belong to a journalist/news organisation or not.

10.2.1 'True' rumours


M @MashableNews [FR: 3, **verified**, **journalist**]
 Stills from eyewitness video show two #CharlieHebdo attackers wearing hoods & black clothing shoot a wounded man
pic.twitter.com/lX8Wys3p5M

Annotation: support: supporting, evidentiality: url-given, certainty: certain,
 see tweet on Twitter


101 retweets
 see tweet on Twitter


 @Raquel75 [FR: 1, **verified**, **journalist**]
 Terrible " @MashableNews: video show two #CharlieHebdo attackers wearing hoods & black clothing shoot a wounded man
pic.twitter.com/2wyBH1qbse"

Annotation: responsetype-vs-source: agreed, certainty: certain, evidentiality: url-given,
 see tweet on Twitter


 @ckozacko [FR: 0, **verified**, **journalist**]
 @Raquel75 @MashableNews and terrible too

Annotation: responsetype-vs-previous: comment, responsetype-vs-source: comment,
 see tweet on Twitter


 @marco_spiro [FR: 0, **verified**, **journalist**]
 @Raquel75 @MashableNews Tantum religio potuit suadere malorum
 see tweet on Twitter

 @Jassalicious [FR: 1, **verified**, **journalist**]
 @MashableNews @mashable the man might have been a wounded police officer as per Tv news coverage

Annotation: responsetype-vs-source: comment,
 see tweet on Twitter


 @AwakeDeborah [FR: 1, **verified**, **journalist**]
 That poor man pictured was a Police Officer :(@MashableNews

Annotation: responsetype-vs-source: comment,
 see tweet on Twitter


 @AwakeDeborah [FR: 1, **verified**, **journalist**]
 Police Officer down in picture, killed by armed terrorists who shouted, "the prophet is avenged" after slaughtering him.
 @MashableNews

Annotation: responsetype-vs-source: comment,
 see tweet on Twitter


Annotation: responsetype-vs-source: comment,
 see tweet on Twitter

 @DeborahFStuart [FR: 0, **verified**, **journalist**]
 @AwakeDeborah @MashableNews so sad


Annotation: responsetype-vs-previous: comment, responsetype-vs-source: comment,
 see tweet on Twitter


 @GsRuba [FR: 1, **verified**, **journalist**]
 Inexplicable evil! " @MashableNews: 2 men wearing hoods & black clothing shoot a wounded man #CharlieHebdo #ParisAttack
pic.twitter.com/dlPpRknrb1

Annotation: responsetype-vs-source: comment,
 see tweet on Twitter

 @Giusguglielmi [FR: 0, **verified**, **journalist**]
 @MashableNews @mashable I hope They are taken alive

Annotation: responsetype-vs-source: comment,
 see tweet on Twitter

 @rastatine1 [FR: 0, **verified**, **journalist**]
 @MashableNews c'est incroyable ce qui nous arrive
 see tweet on Twitter

 @cintamnt [FR: 0, **verified**, **journalist**]
 @MashableNews brain washed fanatics. The amount of fervor needed to commit such a horrific act is unimaginable.

Annotation: responsetype-vs-source: comment,
 see tweet on Twitter


 @AwakeDeborah [FR: 1, **verified**, **journalist**]
 The same way ISIS shouts it. However, Al queda (sp?) is claiming the attack. @Ninas5Nina @MashableNews
 see tweet on Twitter

Figure 10.5: Example of a true rumour.

The short conversation in Figure 10.5 refers to a post releasing a set of photos taken from eyewitness video of the Charlie Hebdo attacks in Paris. Here the validity of the original post is not brought into question and the remaining posts take the form of commentary upon it. Notice how the annotations similarly support the certainty of the claim

of the first, evidence-based post and the subsequent character of the posts as commentary. The origination of the post with a news organization also plays into its truth status here.

Something of interest here is the way it demonstrates relatively simply the point made earlier regarding the extent to which people can self-select in order to respond to prior tweets. The relative embedding of posts in the representation above attempts to capture that extent to which tweets are related. However, another point of interest is how the tweeters involved are aligning to subtly distinct matters of topic, distinctions that are not wholly commensurate with the post relationships themselves. What we can see here are three related but different concerns being addressed by the parties who are tweeting. Some are concerned to assign a moral ascription to the matter overall, e.g. 'terrible'. A second set of posts concern themselves with the identity of the wounded man as a police officer. The third set are focused upon the utterances shouted out by the attackers. A further level of topical richness is present in the way that one of the tweeters, @AwakeDeborah, is actually addressing both the identity of the wounded man and what was shouted out. One post, from @Glusguglielmi, also looks at the prospective actions to be taken to capture the attackers.

The mentions here indicate the post relations to some degree. First of all every post cites @MashableNews, the producer of the original tweet. A couple of others also refer to @Raquel75 who produced the assessment 'terrible', reinforcing this in various ways. Suggestions within the stream that the wounded man is a police officer, however, which (at least temporally) originate with @Jassalicious, make no mention of the original poster. It is taken up (perhaps independently) by @AwakeDeborah and it is her posts that get a mention from @DeborahFStuart with another value assessment: "So sad". @AwakeDeborah's posts are actually interesting also for how they display the scope for Twitter to support extended conversational turns and the way the temporal organization of the stream together with the self-selection norm can result in extensive fragmentation of such turns. Her full turn effectively amounts to "That poor man pictured was a police officer . . . killed by armed terrorists who shouted, "the prophet is avenged" after slaughtering him. . . The same way ISIS shouts it. However, Al queda (sp?) is claiming the attack". However, the latter part of this sits as an isolated and seemingly disconnected post at the end of the stream, undermining its topical coherence in relation to the preceding posts (to which it appears irrelevant) until one sees how it is a continuation of her second post instead. The attempt by users to produce extended turns of this order, and the degree to which Twitter has not yet developed an interface that can properly that can properly support this kind of routine conversational feature, is indicative both of what Twitter may still need to do to evolve into a turn-exchange system like conversation and also of the challenges that still confront us to be able to analyse turns in a sequence. What a user producing tweets may understand as a turn in a sequence, and how it appears as a feature of the Twitter conversational sequence are clearly distinct. In that isolated fragments of a turn may appear out of sequence, this provides for further scope for them to be potentially misunderstood, and that in turn provides for the possibility of such fragments to become a source of misinformation and rumour, regardless of what the originator may have been

trying to accomplish.

The screenshot shows a Twitter thread starting with a tweet from @Independent (verified journalist) about the Charlie Hebdo attacks. The tweet text is: "The last person killed in Charlie Hebdo attacks was Muslim police officer ind.pn/1yE1GTq pic.twitter.com/PJH6tlyrpC". The tweet has 182 retweets. Below it are several replies from other verified journalists and users, including @ivsalmibides_, @andreemurphy, @mr_moog, @Traderunner1, @RanaKBresse, @NelsontheCat01, @nigel2john, @Sadafiqureshi, @PolicemansLot, and @wilkotwig. The replies contain various annotations such as 'support: supporting', 'agreed', 'disagreed', and 'comment', along with references to other tweets or images.

Figure 10.6: Example of a rumour on the Charlie Hebdo killings, where the truth status of the original post is not brought to question.

The example in Figure 10.6 is supplied by way of contrast because here again the topic is photos taken of the Charlie Hebdo killings, once again the truth status of the original post is not brought into question and the originating source is a news organization. However, this time the nature of the original post is called to account in other ways, attending to the appropriateness of making the post in the first place. Annotation of the originating post indicates a very similar status to the previous example, but many of the other posts are not just considered to be comments but to also be engaged in agreement or disagreement with other posts within the stream.

The basic bone of contention here is whether showing a picture of a man at the moment of his death is an insensitive and inappropriate thing to do for a news organisation. Topically, as with the preceding post, this conversation is richer than it might at first seem.

The proposition, first raised by @andreemurphy and subsequently aligned with by 4 other tweets, that the Independent should not have posted the picture, actually becomes a vehicle for articulating a discussion regarding the relationship between Islam and terrorism and the latter posts in the stream are largely addressed more to this topic, motivated by the post initially from @Ranask35 “Islam never Support Terrorist”.

So what we see across both of the Twitter conversations provided here as examples of what might be categorized within the annotation scheme as ‘True’ rumours, is a buried topical richness within what can be broadly classified as comments, that demonstrates how originating posts can be implicative for a wide range of conversational actions that may themselves then be implicative in other ways. One of the features of the Simplest Systematics for Conversation model outlined by [Sacks et al., 1974] is the way conversational turns have a constraining effect upon what subsequent turns within the sequence may look like. What we see happening here in Twitter is an ongoing orientation to topical coherence that seems to carry across such systems, but also, through the vehicle of ready-to-hand and temporally discontinuous self-selection, a means of extending beyond the bounds of what work topical coherence might do within conventional face-to-face conversation. This is significant for how chains of related and unrelated content may unfold and how it may be worked with by others and it provides an indication of another systematic feature within Twitter that may provide for its effectiveness as a vehicle for rumorous content.

10.2.2 ‘False’ rumours

The Twitter stream in Figure 10.7 tackles a prospectively rumorous post from a variety of perspectives, displaying a number of ways in which accountability mechanisms may be visibly brought to bear upon unfolding content of this kind. Ultimately it turns out that the foundation of the post as a ‘false’ rumour hinges upon a confusion of events. The initial post cites the New York Times as saying the Canadian soldier shot in the Ottawa shootings has died. Responses to this initially don’t bring it into question and instead align with content in ways that are similar to the example in Figure 10.5. However, a post then enters the stream saying that Canadian TV is reporting that the soldier is alive. There are then numerous posts aligning with this post, some of which call the original tweeter to account for having posted false information. It is only towards the end of the stream that a post turns up that suggests the possibility that there has been a confusion of events with the death of the soldier referring to an earlier event in Quebec instead. Annotations here are richer capturing the mix of responses with tags that indicate both disagreement with the original post and agreement with subsequent posts, as well as appeals for more information. Interestingly there are subtleties of interchange here that evade capture in the scheme, indicating areas in which there is scope for further work. The critical post that clarifies the source of confusion and provides a point of disambiguation is annotated as a comment, which, within the constraints of the scheme, it clearly is. What is not yet captured in sufficient depth perhaps is the range of conversational work ‘comments’ may

The image shows a screenshot of a Twitter thread with several tweets and their corresponding annotations. The annotations are color-coded: red for certain information and blue for uncertain information. The tweets include:

- @DaveBeninger (FR: 1, verified, journalist)**: BREAKING NEWS: New York Times is reporting the Canadian soldier who was shot has died from their injuries. Heartbreaking. #cdnpoli #abtlg. *Annotation: certainty: certain, evidentiality: source-quoted, support: supporting.*
- @datluka (FR: 0, verified, journalist)**: @DaveBeninger My heart goes out to the family. *Annotation: responseType-vs-source: comment.*
- @ndmartens (FR: 0, verified, journalist)**: @DaveBeninger prayers and thoughts to his family and friends... *Annotation: responseType-vs-source: comment.*
- @MaggieinRDAB (FR: 0, verified, journalist)**: @DaveBeninger Not according to what I've just heard on CTV. *Annotation: certainty: somewhat-certain, evidentiality: no-evidence, responseType-vs-source: disagreed.*
- @teesock (FR: 0, verified, journalist)**: @DaveBeninger @cheryl norrad Ugh, CTV is reporting he's ALIVE. pic.twitter.com/CprCsb3ES9 *Annotation: certainty: certain, evidentiality: source-quoted, responseType-vs-source: disagreed.*
- @shawnynch23 (FR: 0, verified, journalist)**: @DaveBeninger @frednewschaser CTV just said he is being treated and is stable? *Annotation: certainty: uncertain, evidentiality: source-quoted, responseType-vs-source: disagreed.*
- @CharleyPride78 (FR: 0, verified, journalist)**: @DaveBeninger @big_rudo NYT may be wrong, cuz CTV news1 has said that the soldier at Memorial is alive at hospital! *Annotation: certainty: somewhat-certain, evidentiality: source-*
- @loas_la (FR: 0, verified, journalist)**: @DaveBeninger maybe you should get your info a little more locally. *Annotation: responseType-vs-source: comment.*
- @reganohie (FR: 0, verified, journalist)**: @DaveBeninger @bebeasley Careful when reporting on life and death. Let's wait until we really know... when the news is not so chaotic. *Annotation: responseType-vs-source: comment.*
- @Donnajherold (FR: 0, verified, journalist)**: @DaveBeninger @doctorfullerton no report he has died! *Annotation: certainty: uncertain, evidentiality: reasoning, responseType-vs-source: disagreed.*
- @doctorfullerton (FR: 1, verified, journalist)**: @Donnajherold @DaveBeninger New report indicates he is alive but obviously gravely injured. I'm not sure where New York Times got that. *Annotation: certainty: somewhat-certain, evidentiality: source-quoted, responseType-vs-source: disagreed, responseType-vs-previous: agreed.*
- @Donnajherold (FR: 0, verified, journalist)**: @doctorfullerton @DaveBeninger me either. There are reports he is still alive. But not looking good! #Chaveshooting *Annotation: certainty: somewhat-certain, evidentiality: no-evidence, responseType-vs-source: disagreed, responseType-vs-previous: agreed.*
- @NaincaR (FR: 0, verified, journalist)**: @Donnajherold @doctorfullerton @DaveBeninger hi Donna I think we are talking 2 different incidents now the quebec one is what I referred 2 *Annotation: responseType-vs-source: comment, responseType-vs-previous: comment.*
- @Donnajherold (FR: 0, verified, journalist)**: @NaincaR @doctorfullerton @DaveBeninger 77727 *Annotation: certainty: uncertain, evidentiality: no-evidence, responseType-vs-source: appeal-for-more-information.*
- @NaincaR (FR: 0, verified, journalist)**: @Donnajherold @doctorfullerton Canadian news contradicts this *Annotation: certainty: uncertain, evidentiality: no-evidence, responseType-vs-source: disagreed.*
- @ShellaGunnReid (FR: 1, verified, journalist)**: @DaveBeninger @JohnW_MPState... @DaveBeninger *Annotation: responseType-vs-source: comment, responseType-vs-previous: comment.*
- @DaPro3 (FR: 0, verified, journalist)**: @ShellaGunnReid @DaveBeninger perhaps vifidul thinking on my part but still hopeful it is not true. TV still not reporting any CF deaths *Annotation: responseType-vs-source: comment, responseType-vs-previous: comment.*
- @TimothyEWilson (FR: 1, verified, journalist)**: @DaveBeninger @ShellaGunnReid Perhaps wait for Canadian source. *Annotation: certainty: uncertain, evidentiality: no-evidence, responseType-vs-source: appeal-for-more-information.*
- @ShellaGunnReid (FR: 1, verified, journalist)**: @TimothyEWilson twitter.com/JohnW_MPState... @DaveBeninger *Annotation: responseType-vs-source: comment, responseType-vs-previous: comment.*

Figure 10.7: Example of a false rumour.

be seen to accomplish.

The primary cohering feature across this exchange is the mention throughout of the original tweeter, @DaveBeninger. Groups of tweets within the exchange, however, then cohere around a range of other mentions, not all of whom are even visible within the stream, possibly because they simply retweeted the posts of others, e.g.: @cheryl norrad; @frednewschaser; @big_rudo; @bebeasley; and so on. A matter of potential importance here is the kinds of considerations being brought to bear by people when they use mentions within conversational streams like this. A question open to further exploration, for instance, is the extent to which people mention people they directly follow, even if they only received a retweet, rather than simply tagging their response with the names of originators within the overall conversation. It is possible that mentions like this may serve as a way of 'doing politeness', providing an acknowledgement of their own source and showing an orientation towards accountability for this kind of work. If tweeters systematically choose to cite the person they follow as an originator within a conversational stream this provides a prospective future mechanism for unraveling how the chains of communication within the spread of a rumour may have unfolded.

Another small feature of interest within this particular example is the way in which disalignment and dispute of the original tweet gets marked out within the textual realization of the response. This echoes observations in the conversation analytic literature of how speakers will mark out dispreferred responses in conversation, such as disagreement, in unique ways that provide for the seaability of the up-and-coming response and giving

the original speaker an opportunity to therefore engage in repair. The most classic marker here is a pause before reponse, or the extended use of non-linguistic utterances such as 'erm', or 'um'. Because Twitter is neither face-to-face nor temporally continuous in its organization these kinds of devices, attuned tightly to the systematic concern discussed above of providing for the avoidance of conversational gaps and overlaps, are of little effect. However, if we look at the posts above we see some interesting small details that are worth pointing out. Thus @teesock's first post that indicates the original post is open to questions contains two points of note: 1) the prefacing of the post with 'Ugh', which may not amount to a delay of reponse in conventional terms but does nonetheless seem to do work as a dispreference marker, testifying perhaps to the legacy of an existing conversation apparatus that informs how people understand they may go about doing things like this; 2) the capitalization of the final word 'ALIVE', which also serves to mark out the specific detail being contested. We then see the following post from @shawlynch23 closing the post with a question mark which subtly marks out the manner in which the disagreement is being done: this contesting position is not being proposed as certain; adding question intonation to the concluding word of the text also alludes to how this remark might be delivered in actual face-to-face conversation, once again indicating that tweeters themselves may be attuned to a situated conversation base for how they construct some of their responses. The subsequent post from @CharleyPride78 also does some of the same kind work by concluding the text with an exclamation mark. So none of these immediately contesting posts engage in simple bold counter-assertion to indicate that a post may be false. Instead they draw on the mechanics of ordinary everyday conversation to mark out a manner of dispute that testifies to a recognition of the fact that disagreement is somehow also dispreferred as a conversational act.

10.2.3 'Unverified' Rumours

In Figure 10.8 we see a relatively straightforward interchange unfolding around a post that has circulated an unverified report from the Daily Mail that Vladimir Putin might be missing. The subsequent posts are all of the order of comments upon this, delivering a range of perspectives including a video-based joke of kinds that points to a clip on YouTube. The annotations capture accurately the potentially dubious character of the claim made in the originating post and the status of the rest of the conversation as commentary.

What is interesting here is the way in which the subsequent comments do not directly express skepticism regarding the status of the original post but rather offer up a series of speculations regarding what an explanation might be if the rumour turns out to be true. Nearly all of the posts indicate a subtle orientation to the potentially dubious character of the rumour by providing responses that are readably humorous or tongue in cheek in some way. This is something for which it is extremely hard to annotate in any rigorous fashion but which, in its articulation, indicates how an original post may be being oriented to by its recipients. A further important feature here relates to how the original post is itself delivered which is also articulated in potentially humorous terms. Conversationally-speaking

@PatOndabak [FR: 1, **verified, journalist**]
Was Vladimir Putin neutralized by an internal coup? Or maybe he's vacationing in Harper's closet: linkis.com/www.dailymail...
[#LookingForNarnia](#)

Annotation: support: supporting, evidentiality: url-given, certainty: somewhat-certain,
7 retweets
see tweet on Twitter

@lordWatcher85 [FR: 0, **verified, journalist**]
@PatOndabak I'm going with neutralized and their entire gov aQuake & pissing selves w fear. Night of long knives gonna take months :o)))

Annotation: responsetype-vs-source: comment,
1 retweets
see tweet on Twitter

@blondewbraintoo [FR: 1, **verified, journalist**]
@PatOndabak or maybe he just needs a break from all the shit going on in this world.. I know I do!

Annotation: responsetype-vs-source: comment,
1 retweets
see tweet on Twitter

@UncleRee1 [FR: 1, **verified, journalist**]
@PatOndabak It is disconcerting that the man with the codes to the worlds largest nuclear arsenal is missing. I suspect an unhappy oligarch.

Annotation: responsetype-vs-source: comment,
see tweet on Twitter

@kelownascott [FR: 0, **verified, journalist**]
@PatOndabak
Putin has gone 1 better
Behold the Russian secure office
Actor portrayed even
youtube.com/watch?v=pi6bRt...

Annotation: responsetype-vs-source: comment,
see tweet on Twitter

@surfeitndearth [FR: 0, **verified, journalist**]
@PatOndabak Maybe, just maybe... he's huddling with best minds/top advisors about how to prevent [#USA](#) [#JK](#) [#NATO](#) [#EU](#) [#Canada](#) threat of WWII ?

Annotation: responsetype-vs-source: comment,

Figure 10.8: Example of an unverified rumour.

seebly humorous or tongue-in-cheek utterances constrain what accountable responses might look like. Typically, to take a readably humorous proposition as a serious statement and to produce an utterance that has visibly analysed it that way is, in ordinary everyday conversation, subject to a variety of reasoned responses that effectively serve to close the inappropriate character of the response down. These can be by means of: a) repair – expressly explaining the 'joke' so that the other party can see what was intended; b) subjecting the inappropriate response to a teasing response of some kind in its own right, which serves much the same function as the preceding case by making it visible that the original utterance was taken the wrong way; or c) by reasoning in various overt or less overt ways that the respondent 'has no sense of humour', and thus edging them out of the conversation by not treating the inappropriate utterance as implicative for their own subsequent utterances in any way. In the above sequence we can see that the contributors to the stream have all recognized and worked within the constraint of the way the original post delivered as prospectively tongue-in-cheek. A question that arises here, and that is of moment for the potential spread of rumours, is the degree to which the kinds of repair and manage strategies for handling inappropriate readings of original utterances in conversation outlined above, may exist within the organization of conversational streams on Twitter. Without local management of this kind, scurrilously rumorously tweets may well be open to having wrong readings of them open to propagation without control.

10.2.4 Speculation

@scATX [FR: 1, verified, journalist]
 #MikeBrown wanted to walk down the street. And he is dead. And the #Ferguson PD is using his silence in death to make him a criminal.

Annotation: certainty: certain, evidentiality: no-evidence, support: supporting,
 108 retweets
 see tweet on Twitter

@mcbyrne [FR: 0, verified, journalist]
 @scATX it's sound like they may have been racially profiling him as a suspect

Annotation: responsetype-vs-source: comment,
 1 retweets
 see tweet on Twitter

@Moneyman2626 [FR: 0, verified, journalist]
 @scATX You are supposed to not be biased when you're a journalist. Glad you can come up with that decision without knowing all the facts

Annotation: certainty: uncertain, evidentiality: no-evidence, responsetype-vs-source: disagreed,
 see tweet on Twitter

@DebraWinters28 [FR: 0, verified, journalist]
 @scATX If they were criminals why did the police not arrest his friend right away? They shot one robber & let the other go #MikeBrown

Annotation: certainty: somewhat-certain, evidentiality: no-evidence, responsetype-vs-source: appeal-for-more-information,
 see tweet on Twitter

@adjordan [FR: 1, verified, journalist]
 @scATX @taiping2 a story that continues to repeat itself, over and over and over in America. #Ferguson

Annotation: responsetype-vs-source: comment,
 see tweet on Twitter

@WebsterGilley [FR: 0, verified, journalist]
 @scATX no he was walking IN the street big difference

Annotation: certainty: certain, evidentiality: reasoning,
 responsetype-vs-source: disagreed,
 see tweet on Twitter

Figure 10.9: Example of an speculative rumour.

In the conversation in Figure 10.9 that tackles the topic of the shooting of the black teenager Mike Brown by police in Ferguson, we see an original post that alludes to speculation regarding the status of Mike Brown as a criminal or otherwise and the motives of the police regarding how they are handling his death. Despite its brevity the unfolding work going on in this conversation is more complex and this, too, is captured by the annotations, which recognize the absence of evidence attaching to the claims present in the originating post and presence of both disagreement with the post and with the need for more evidence in the subsequent stream.

What is interesting here is the range of work that can be going on within something that might be broadly characterized as disagreement. It is worth tracking this through a little:

So, in the opening post from @scATX we can see a series of relatively bald statements that, in the absence of supporting materials, might be read as the expression of an opinion that implies a particular, speculative view regarding the motives of the Ferguson Police

Department in the matter. All of the subsequent posts are addressed to this initial post, marking out their relationship to it with a preliminary mention of @scATX. One post, from @adjordan also mentions another party, @taiping2, whose posts are not visible in the stream (see the discussion of mentions like this in section 6.2.2 above). The first response post from @mcbyrne effectively aligns with the original post and adds a further layer of speculation, i.e. that the police 'may have been racially profiling him as a suspect'. However, the next post from @Moneyman2626 is quite distinct. Here @scATX is effectively called to account for the original post, the implication being that @scATX has proceeded in an inappropriate way for a journalist (which, importantly, indicates specific knowledge regarding the original source on the part of this respondent). Note also how the calling to account here is managed: it is not a direct case of 'what do you mean?', or 'why would you say that?', etc. Instead there is first of all a statement of how journalists should behave and then a readably sarcastic description of how @scATX is seen to have behaved. This an obvious case of disagreement but, in the work it accomplishes as a calling to account, it is also much more than that, demonstrating the extent to which Twitter offers similar mechanisms for doing this kind of work, despite the organisational constraints within which tweeters have to operate. The subsequent post from @DebraWinters28 is doing something else again. Here she tacitly works with @scATX's proposition that the police have decided to dress Mike Brown up as a criminal and uses it to question that reasoning the police were using if that was the case, finding inconsistencies in their behaviour that may be read equally as alignment with @scATX's point of view, pointing to flaws in police practice instead, or as grounds for questioning @scATX's proposition by finding it inconsistent with how the police behaved. The next post from @adjordan is more clearly aligned with the original post and it uses @scATX's proposition as a vehicle to index a story of inappropriate police behaviour across America, which is once again more than just simple alignment and nothing else. The final post from @WebsteGilley takes a different line and reformulates the original post slightly to make it readable in a different way: it was not Mike Brown's movements that led to his death but, by implication, the fact he was there at all. This post is particularly problematic for annotation because it is not quite just a case of disagreement, In some ways it is actually aligned with the original post. However, it uses a routine conversational practice of reformulation to transform the original post to speak to a different set of concerns to the ones that might have originally motivated it.

What we see, then, in the detail of how this conversation unfolds, is a series of quite distinct accomplishments that are not easy to capture in generic terms, demonstrating how, as one moves into conversational domains where there is larger potential for a range of possible responses, such as in the case of speculation and controversy, there may be a need to develop further ways of handling the different implications of the different kinds of response that moves beyond just questions of agreement or disagreement with an original post.

10.2.5 Controversy

The figure shows a screenshot of a Twitter thread. The main tweet is from Mashable, reporting on the Germanwings crash. Below it are several replies from various users, each with a corresponding annotation. The annotations capture the presence of potentially unverified elements in the originating post and the variety of responses it generates. The controversy is centered on the choice of words used to describe the event, specifically the use of '911' versus 'Mayday'.

Annotations include:

- @mashable [FR: 4, verified, journalist]**: Latest on #Germanwings crash: Pilots signaled 911 before dropping out of midair; airline CEO calls this a "dark day." on.mash.to/1EE83U
- @dsabar [FR: 0, verified, journalist]**: @mashable Signalled 911? Called 'Mayday' would be more appropriate, factual reporting...
- @latludehopper [FR: 0, verified, journalist]**: @mashable signaled 911?
- @Duke_mochuck [FR: 1, verified, journalist]**: @mashable What does that even mean? Why would European pilots say 911 - an American emergency phone number? Nothing in article about that.
- @Mark_R_Woods [FR: 1, verified, journalist]**: @mashable you might want to change the use of 911 in this context.
- @zywo [FR: 1, verified, journalist]**: *@mashable: Latest on #Germanwings crash: Pilots signaled 911 before dropping out of midair' dropping out
- @davidjindlay [FR: 0, verified, journalist]**: @mashable #germanwings really? pilots signaled 911? So they got on a satellite phone and called a US emergency line?
- @dobykacyc! [FR: 0, verified, journalist]**: @mashable yessss!
- @thebavery [FR: 0, verified, journalist]**: @mashable Did you mean 'mayday' here or "...for emergency"?
- @pushreply [FR: 0, verified, journalist]**: @mashable didn't find '911' in all stories and updates, international and local german. Please don't speculate and spread lies
- @brianmoreau [FR: 0, verified, journalist]**: @mashable didn't find '911' in all stories and updates, international and local german. Please don't speculate and spread lies
- @brianmoreau [FR: 0, verified, journalist]**: @mashable makes reference to 911 RE #Germanwings NOTHING BUT LIES at the cost of LIVES, BOYCOT website
- @brianmoreau [FR: 0, verified, journalist]**: @mashable en.wikipedia.org/wiki/Pete_Cash... the man behind the lies
- @brianmoreau [FR: 0, verified, journalist]**: @mashable His Facebook, facebook.com/petecashmore Address coming shortly
- @Nilltejunder [FR: 1, verified, journalist]**: @mashable:Latest on #Germanwings crash: Pilots signaled 911 b4dropping out of midair.airlineCEO calls this a'dark day'on.mash.to/1EE83U
- @rivsony [FR: 1, verified, journalist]**: Si c'est comme ça que les français vont faire des recherches, c'est sûr qu'ils ne vont rien trouver @mashable

Figure 10.10: Example of a controversial rumour.

The example in Figure 10.10 presents a conversation that unfolds regarding the final moments of the German Wings plane before it crashed in the French Alps. The annotations capture both the presence of potentially unverified elements in the originating post and the variety of responses it generates. What is interesting here is how the focus of the controversy is not so much the accuracy or otherwise of the basic claim that a distress signal came from the plane before it crashed, but rather the choice of words to gloss that which are used in the original Mashable posting. Indeed, all of the subsequent posts are in some way directed to the choice of signaling '911' to describe the claimed event, taking Mashable to task for having expressed it in this way.

The importance of this example is in how it demonstrates a need to be attentive to the grounds of disagreement with an originating post. The bulk of the controversy here revolves around the fact that Mashable, clearly oriented to an American audience, has use a specific Americanism to describe the production of a distress call, i.e. 'calling 911', though it should be noted that one of the posts in the stream also questions that description 'dropping out of midair'. If, as several of the posts suggest, a term such as calling 'Mayday' had been used, clearly the grounds of controversy here would have been quite distinct. One can call a tweeter to account for the appropriateness or otherwise of using a specific term quite readily, as the flurry of posts here demonstrates. A calling to account for the post if it had used the term 'Mayday' would have needed to be more concerned

with the accuracy of the claim itself, not the embedded descriptors. This is an altogether different order of work. Having posted this as a reasonable claim the questioning of its accuracy has to itself present grounds for why it might want to call the claim into question (see the discussion of the Ottawa shooting report in section 6.2.2 above). Failure to present a reason for why a claim is considered dubious is to allow for the possibility that the questioning might itself be called to account, with just saying you don't believe it not being quite enough. What this underscores is that, whilst controversy about specific details might take the form visible in the example above, controversy directed to the actual claims present in posts needs to be systematically organized around the form of claims and counterclaims to meet the demands of ordinary conversational practice. And it would seem to be the case that this organizational principle is as relevant in Twitter as it is in other language based turn-taking systems where the production of a visible claim is possible in the first place.

10.2.6 Agreement

The figure shows a vertical thread of tweets on Twitter. The top tweet is from @9NewsSyd, a verified journalist, stating: "#BREAKING: Police have confirmed that the #SydneySiege is over. #9News". Below this tweet is an annotation: "Annotation: certainty: somewhat-certain, support: supporting, evidentiality: witnessed, 677 retweets".

Below the first tweet are several replies. One reply from @TheBasedNoel_ [FR: 1, verified, journalist] says: "@@9NewsSyd: #BREAKING: Police have confirmed that the #SydneySiege is over. #9News" thank God bY™CEbY™bY". An annotation for this reply reads: "Annotation: responseType-vs-source: comment, see tweet on Twitter".

Another reply from @Nikic308 [FR: 0, verified, journalist] says: "@9NewsSyd good". An annotation for this reply reads: "Annotation: responseType-vs-source: comment, see tweet on Twitter".

Further down, a tweet from @WilliamBurge98 [FR: 0, verified, journalist] says: "@@9NewsSyd: #BREAKING: Police have confirmed that the #SydneySiege is over. #9News" Thank god... Shoutout to all the police that saved us". An annotation for this tweet reads: "Annotation: responseType-vs-source: comment, see tweet on Twitter".

Other tweets in the thread include:

- @justmattaye [FR: 0, verified, journalist]: "@@9NewsSyd: #BREAKING: Police have confirmed that the #SydneySiege is over. #9News" so glad to hear, this was unreal! Annotation: responseType-vs-source: comment, 1 retweets.
- @MabelleXO [FR: 1, verified, journalist]: "@@9NewsSyd: #BREAKING: Police have confirmed that the #SydneySiege is over. #9News" bY™ Annotation: responseType-vs-source: comment, 2 retweets.
- @7WickedWitch [FR: 0, verified, journalist]: "@9NewsSyd Por Fin!" Annotation: responseType-vs-source: comment, 3 retweets.
- @niellegreivty [FR: 2, verified, journalist]: "@@9NewsSyd: #BREAKING: Police have confirmed that the #SydneySiege is over. #9News" best news Annotation: responseType-vs-source: comment, 3 retweets.
- @vampiremoneys [FR: 0, verified, journalist]: "@9NewsSyd YES" Annotation: responseType-vs-source: comment, see tweet on Twitter.
- @Shar16 [FR: 1, verified, journalist]: "@@9NewsSyd: #BREAKING: Police have confirmed that the #SydneySiege is over. #9News" bY™bY™ Annotation: responseType-vs-source: comment, 2 retweets.
- @harryfmusic [FR: 1, verified, journalist]: "@@9NewsSyd: #BREAKING: Police have confirmed that the #SydneySiege is over. #9News" bY™bY™ good job australia, hope they all are okay Annotation: responseType-vs-source: comment, see tweet on Twitter.
- @stilyinwhore_ [FR: 0, verified, journalist]: "@@9NewsSyd: #BREAKING: Police have confirmed that the #SydneySiege is over. #9News" I would not only like to thank God, but those policemen Annotation: responseType-vs-source: comment, 1 retweets.
- @alluiveharry [FR: 2, verified, journalist]: "@@9NewsSyd: #BREAKING: Police have confirmed that the #SydneySiege is over. #9News" thank god thank god thank you thank you thank you Annotation: responseType-vs-source: comment, see tweet on Twitter.
- @sunnywaggekites [FR: 1, verified, journalist]: "@9NewsSyd THANKS GOD" Annotation: responseType-vs-source: comment, see tweet on Twitter.
- @amy_15666 [FR: 0, verified, journalist]: "@9NewsSyd THANK GOODNESS" Annotation: responseType-vs-source: comment, see tweet on Twitter.

Figure 10.11: Example of an agreement rumour.

In the final example in Figure 10.11 we see a series of posts relating to the Sydney siege and the news that the siege has ended. The initial annotations capture the fact that the support for the initial post is based upon a series of hashtags rather than urls or other evidence, and the post is actually assigned a status that is only 'somewhat certain'.

Nonetheless, subsequent annotations recognize the way in which the rest of the conversation amounts to a set of aligning comments on this initial post with varying indicators of relief and commendation for the Australian police. All of which adds up to a strong body of agreement even though it is not quite characterized that way.

The problem here is one of identifying just what is being aligned with exactly. The Sydney News source here that originates the claim that the siege has ended does not take a specific moral stance with regard to the news. So what one gets here is not strong agreement with the claim the siege is over (and we have already discussed above how the work of contesting such a claim would look quite different anyway). Instead there is a strong body of alignment with regard to what an appropriate moral response to such news should look like. The issue to be confronted here is that this kind of alignment, although distinct from the original post, is fateful for how an unfolding rumour may get handled. We can see this in the example in Figure 10.7 where what might be termed 'appropriate moral responses' to the news of the death of a soldier are already beginning to bubble up when they are cut across by the production of a direct counterclaim to the news. It is hard to produce an 'appropriate moral response' without also providing some kind of indicator as to the thing you are responding in that way to. Thus these kinds of ordinary conversational productions of alignment and agreed forms of response are tremendously powerful for simultaneously promoting the original claim in order to evidence what is being responded to. Thus it is no coincidence that what one sees in the example above are multiple reiterations of the original source tweet as a component part of the response.

We have emphasized above the importance of attending to the three turn structure and the need to look at not just source tweets but how they are handled in turn by other responses on the timeline. What we see in the specific kinds of cases of agreement presented here is a need to also systematically consider how the response tweets may themselves have important implications for what might be classed as agreement, speculation, controversy, and so on. Looking back one again to Example 3 above it can be seen that the annotation scheme has a built in mechanism for dealing with this kind of concern because it provides annotators with a means of annotating regarding the nature of the response not just in relation to the source tweet but also in relation to the previous tweet in the conversational stream.

10.3 Conclusion

In this chapter we have sought to present the way in which we have turned to the social scientific disciplines of conversation analysis and ethnomethodology in order to extend beyond the initial annotation scheme and to begin to enrich it with a deeper body of understanding of real-world social practice and interactional methodologies. We have also outlined how the accomplishment of this turns upon moving beyond just using conversation analysis as a frame and involves instead taking microblogging and the use of Twitter as a discrete domain of practice that requires analysis in its own terms. In order

to explicate this further we have looked systematically at a range of significant objects of interest in the conversation analytic literature and how the handling of these needs to be reconfigured in order to make it speak properly to microblogging as an organizational phenomenon in its own right.

Finally, we have taken a small set of examples of already annotated rumours in the corpus in order to examine how some parts of the approach we have been advocating may be brought to bear in order to further enrich our understanding of just what might be going on over the course of the production and dissemination of such tweets. Clearly such analysis can be extended much further and there is a need to further refine our understanding of the organizational characteristics of Twitter-use, but already it can be seen that valuable progress in this direction has been made.

Chapter 11

Discussion

This document describes our efforts for developing an annotation scheme for rumours spread on social media, putting it into practice for the annotation of a large-scale dataset of social media rumours, as well as performing a qualitative analysis of these annotated rumours. The annotation scheme has been developed in the context of the PHEME project's Work Package 2, and it enables the annotation of conversation aspects that occur around rumours in social media. Having tested and validated this annotation scheme through an iterative process, including experienced users on site at UWAR's premises, as well as crowdsourced annotators recruited online, we have applied it to a large-scale dataset. The dataset used as input at this point has been put together in collaboration with SWI and ATOS for Work Package 8. This dataset of rumours and non-rumours has been enriched for our purposes of analysing conversations around rumours, and expanded with more content from other media, languages, and additional metadata. The application of the annotation scheme to the resulting dataset has produced an annotated corpus of 330 Twitter threads in English and German, including overall 4,842 tweets.

The qualitative analysis performed on a small subset of this annotated corpus allowed us to examine how some parts of the approach we have been advocating may be brought to bear in order to further enrich our understanding of just what might be going on over the course of the production and dissemination of such tweets.

The dataset produced in this Work Package and presented in this deliverable will be publicly released in month 24. At the time of delivering this document, we release a small sample of it, which includes 8 threads in English and 2 in German. This dataset is an expanded version of that released in WP8, providing scheme-based annotation for a subset of 330 threads, as well as information flows and media links.

Bibliography

- [Antaki, 2000] Antaki, C. (2000). Two rhetorical uses of the description ‘chat’. *M/C: a Journal of Media and Culture*, 3(4).
- [Bergmann, 1993] Bergmann, J. R. (1993). *Discreet indiscretions: The social organization of gossip*. Transaction Publishers.
- [Cheng et al., 2015] Cheng, J., Teevan, J., Iqbal, S. T., and Bernstein, M. S. (2015). Break it down: A comparison of macro- and microtasks. In *Proceedings of CHI*.
- [Clifton, 2009] Clifton, J. (2009). A membership categorization analysis of the waco siege: Perpetrator-victim identity as a moral discrepancy device for ‘doing’ subversion. *Sociological Research Online*, 14(5):8.
- [Coulter, 1979] Coulter, J. (1979). Beliefs and practical understanding. *Everyday Language. Studies in Ethnomethodology*. New York: Irvington Publishers.
- [De Choudhury et al., 2012] De Choudhury, M., Diakopoulos, N., and Naaman, M. (2012). Unfolding the event landscape on twitter: classification and exploration of user categories. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 241–244. ACM.
- [de Marneffe et al., 2011] de Marneffe, M., Manning, C. D., and Potts, C. (2011). Veridicality and utterance understanding. In *Semantic Computing (ICSC), 2011 Fifth IEEE International Conference on*, pages 430–437. IEEE.
- [DiFonzo and Bordia, 2007] DiFonzo, N. and Bordia, P. (2007). Rumor, gossip and urban legends. *Diogenes*, 54(1):19–35.
- [Durkheim et al., 1938] Durkheim, E., CATLIN, S. G. E. G., and SOLOVAY, S. A. (1938). *The Rules of Sociological Method... Translated by Sarah A. Solovay and John H. Mueller, and Edited by George EG Catlin*.
- [Finin et al., 2010] Finin, T., Murnane, W., Karandikar, A., Keller, N., Martineau, J., and Dredze, M. (2010). Annotating named entities in twitter data with crowdsourcing. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk*, pages 80–88. Association for Computational Linguistics.

- [Friggeri et al., 2014] Friggeri, A., Adamic, L. A., Eckles, D., and Cheng, J. (2014). Rumor cascades. In *ICWSM*.
- [Garfinkel, 1967] Garfinkel, H. (1967). Studies in ethnomethodology.
- [Garfinkel, 1991] Garfinkel, H. (1991). Respecification: Evidence for locally produced, naturally accountable phenomena of order, logic, reason, meaning, method, etc. in and as of the essential haecceity of immortal ordinary society (i)—an announcement of studies. *Ethnomethodology and the human sciences*, pages 10–19.
- [Goodwin, 1980] Goodwin, M. H. (1980). he-said-she-said: formal cultural procedures for the construction of a gossip dispute activity. *American Ethnologist*, 7(4):674–695.
- [Guerin and Miyazaki, 2006] Guerin, B. and Miyazaki, Y. (2006). Analyzing rumors, gossip, and urban legends through their conversational properties. *Psychological Record*, 56(1).
- [Hannak et al., 2014] Hannak, A., Margolin, D., Keegan, B., and Weber, I. (2014). Get back! you don't know me like that: The social mediation of fact checking interventions in twitter conversations. In *ICWSM*.
- [Harper, 1994] Harper, R. (1994). Radicalism, beliefs and hidden agendas. *Computer Supported Cooperative Work (CSCW)*, 3(1):43–46.
- [Heritage et al., 2001] Heritage, J., Boyd, E., and Kleinman, L. (2001). Subverting criteria: the role of precedent in decisions to finance surgery. *Sociology of Health & Illness*, 23(5):701–728.
- [Jalbert, 1989] Jalbert, P. L. (1989). Categorization and beliefs: News accounts of haitian and cuban refugees. *The interactional order: New directions in the study of social order*, pages 231–248.
- [Koohang and Weiss, 2003] Koohang, A. and Weiss, E. (2003). Misinformation: toward creating a prevention framework. *Information Science*.
- [Lee, 1984] Lee, J. (1984). Innocent victims and evil-doers. In *Women's Studies International Forum*, volume 7, pages 69–73. Elsevier.
- [Lotan et al., 2011] Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., et al. (2011). The arab spring— the revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions. *International journal of communication*, 5:31.
- [Meehan, 1989] Meehan, A. J. (1989). Assessing the “police-worthiness” of citizen's complaints to the police: accountability and the negotiation of “facts”. *The interactional order: New directions in the study of social order*, pages 116–140.

- [Mellinger, 1992] Mellinger, W. M. (1992). "accomplishing fact in police" dispatch packages": An analysis of the situated construction of an organizational record. In *Perspectives on social problems*, volume 4, pages 47–72.
- [Navarro, 2001] Navarro, G. (2001). A guided tour to approximate string matching. *ACM computing surveys (CSUR)*, 33(1):31–88.
- [Parker and O'Reilly, 2012] Parker, N. and O'Reilly, M. (2012). 'gossiping' as a social action in family therapy: The pseudo-absence and pseudo-presence of children. *Discourse Studies*, 14(4):457–475.
- [Paul et al., 2011] Paul, S. A., Hong, L., and Chi, E. (2011). What is a question? crowd-sourcing tweet categorization. *CHI 2011*.
- [Procter et al., 2013a] Procter, R., Housley, W., Williams, M., Edwards, A., Burnap, P., Morgan, J., Rana, O., Klein, E., Taylor, M., Voss, A., et al. (2013a). Enabling social media research through citizen social science. In *ECSCW 2013 Adjunct Proceedings*.
- [Procter et al., 2013b] Procter, R., Vis, F., and Voss, A. (2013b). Reading the riots on twitter: methodological innovation for the analysis of big data. *International Journal of Social Research Methodology*, 16(3):197–214.
- [Pustejovsky et al., 2003] Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Sundheim, B., Day, D., Ferro, L., et al. (2003). The timebank corpus. In *Corpus linguistics*, volume 2003, page 40.
- [Qazvinian et al., 2011] Qazvinian, V., Rosengren, E., Radev, D. R., and Mei, Q. (2011). Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1589–1599. Association for Computational Linguistics.
- [Rapley, 1998] Rapley, M. (1998). 'just an ordinary australian': Self-categorization and the discursive construction of facticity in 'new racist' political rhetoric. *British Journal of Social Psychology*, 37(3):325–344.
- [Ritter et al., 2010] Ritter, A., Cherry, C., and Dolan, B. (2010). Unsupervised modeling of twitter conversations. In *Proc of NAACL*.
- [Rosnow and Foster, 2005] Rosnow, R. L. and Foster, E. K. (2005). Rumor and gossip research. *Psychological Science Agenda*, 19(4).
- [Sacks, 1978] Sacks, H. (1978). Some technical considerations of a dirty joke. *Studies in the organization of conversational interaction*, pages 249–270.
- [Sacks, 1979] Sacks, H. (1979). Hotrodder: A revolutionary category. *Everyday language: Studies in ethnomethodology*, pages 7–14.

- [Sacks, 1995] Sacks, H. (1995). *Lectures on conversation*, volume 1. Blackwell Publishing.
- [Sacks et al., 1974] Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, pages 696–735.
- [Saurí and Pustejovsky, 2009] Saurí, R. and Pustejovsky, J. (2009). Factbank: A corpus annotated with event factuality. *Language resources and evaluation*, 43(3):227–268.
- [Saurí and Pustejovsky, 2012] Saurí, R. and Pustejovsky, J. (2012). Are you sure that this happened? assessing the factuality degree of events in text. *Computational Linguistics*, 38(2):261–299.
- [Sidnell, 2011] Sidnell, J. (2011). 6 the epistemics of make-believe. *The morality of knowledge in conversation*, 29:131.
- [Smith, 1978] Smith, D. E. (1978). K is mentally ill'the anatomy of a factual account. *Sociology*, 12(1):23–53.
- [Soni et al., 2014] Soni, S., Mitra, T., Gilbert, E., and Eisenstein, J. (2014). Modeling factuality judgments in social media text. In *ACL*.
- [Vlachos and Riedel, 2014] Vlachos, A. and Riedel, S. (2014). Fact checking: Task definition and dataset construction. *ACL 2014*, page 18.
- [Walton, 2010] Walton, D. (2010). Types of dialogue and burdens of proof. In *COMMA*, pages 13–24.
- [Wong-Sak-Hoi, 2015] Wong-Sak-Hoi, G. (2015). D8.2 annotated corpus of newsworthy rumours. *PHEME deliverable*.
- [Wooffitt, 1992] Wooffitt, R. (1992). *Telling tales of the unexpected: The organization of factual discourse*. Rowman & Littlefield.
- [Zubiaga and Ji, 2014] Zubiaga, A. and Ji, H. (2014). Tweet, but verify: Epistemic study of information verification on twitter. *Social Network Analysis and Mining*.
- [Zubiaga et al., 2015] Zubiaga, A., Liakata, M., Procter, R., Bontcheva, K., and Tolmie, P. (2015). Crowdsourcing the annotation of rumourous conversations in social media. In *Proceedings of the 24th International Conference on World Wide Web Companion*, pages 347–353. International World Wide Web Conferences Steering Committee.
- [Zubiaga et al., 2014] Zubiaga, A., Tolmie, P., Liakata, M., and Procter, R. (2014). D2.1 development of an annotation scheme for social media rumours. *PHEME deliverable*.