Give me your Intentions, I'll Predict Our Actions: A Two-level Classification of Speech Acts for Crisis Management in Social Media

Enzo Laurenti, Nils Bourgon, Farah Benamara, Alda Mari, Véronique Moriceau, Camille Courgeon

Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 4333–4343
Marseille, 20-25 June 2022

Give me your Intentions, I'll Predict Our Actions: A Two-level Classification of Speech Acts for Crisis Management in Social Media

Enzo Laurenti[1], **Nils Bourgon**[2], **Farah Benamara**[2], **Alda Mari**[1], **Véronique Moriceau**[2], **Camille Courgeon**[1]
[1] IJN, CNRS/ENS/EHESS/PSL University {firstname.lastname@ens.fr}
[2] IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France {firstname.lastname@irit.fr}

Abstract

Discovered by (Austin,1962) and extensively promoted by (Searle, 1975), speech acts (SA) have been the object of extensive discussion in the philosophical and the linguistic literature, as well as in computational linguistics where the detection of SA have shown to be an important step in many down stream NLP applications. In this paper, we attempt to measure for the first time the role of SA on urgency detection in tweets, focusing on natural disasters. Indeed, SA are particularly relevant to identify intentions, desires, plans and preferences towards action, providing therefore actionable information that will help to set priorities for the human teams and decide appropriate rescue actions. To this end, we come up here with four main contributions: (1) A two-layer annotation scheme of SA both at the tweet and subtweet levels, (2) A new French dataset of 6,669 tweets annotated for both urgency and SA, (3) An in-depth analysis of the annotation campaign, highlighting the correlation between SA and urgency categories, and (4) A set of deep learning experiments to detect SA in a crisis corpus. Our results show that SA are correlated with urgency which is a first important step towards SA-aware NLP-based crisis management on social media.

Keywords: Speech acts, Crisis events, Social Media

1. Introduction

The use of social networks is pervasive in our daily life. All areas are concerned, including civil security and crisis management. Recently, Twitter has been widely used to generate valuable information in crisis situations, showing that traditional means of communication between population and rescue departments (e.g., phone calls) are clearly suboptimal (Vieweg et al., 2014; Olteanu et al., 2015). For example, more than 20 million tweets were posted during the superstorm Sandy in 2012 (Castillo, 2016) and the hashtag #NotreDame relatif to the the Notre Dame fire that occurred in France has been the most used in Twitter in 2019.[1]

One crucial aspect of tweets posted during crisis events pertains to the fact that people express their intentions, desires, plans, goals and preferences towards action, providing therefore actionable information that will help to set priorities for the human teams and decide appropriate rescue actions. For example, in the tweet (1a),[2] the writer publicly expresses an explicit commitment to provide help after the Irma hurricane tragedy, using an explicit action verb ("*to help*") which is under the scope of an explicit attitude verb ("*want*"). (1b) on the other hand expresses an intention to complain about the absence of assistance without using any explicit intent keywords. Intention to advise, evacuate (cf. (1c)) are other types of actions expressed in crisis situations.

[1] https://blog.twitter.com/en_us/topics/insights/2019/ThisHappened-in-2019

[2] These are examples taken from our French corpus translated into English.

It is important to note that such useful messages do not always require an urgent and rapid action from rescue teams: messages like (1c), about affected people, or infrastructure damages can be seen as more urgent compared to other types of intention to act (cf. (1a-b)).

(1) a. #Irma Hurricane: "I want to go there to help."
b. Irma hurricane: where is disaster assistance one month later?
c. Emergency heritage at Bordeaux. After the flood, the archaeology lab is looking for volunteers to evacuate collections.

Our work focuses on the impact of speech acts on emergency detection during crises. Discovered by (Austin, 1962) and extensively promoted by (Searle, 1975), speech acts (henceforth SA) have been the object of extensive discussion in the philosophical and the linguistic literature ((Hamblin, 1970; Brandom, 1994; Sadock, 2004; Asher and Lascarides, 2008; Portner, 2018) to mention just a few). According to the Austinian initial view, SA are to achieve action rather than conveying information. When uttering *I now pronounce you man and wife*, the priest accomplishes the action of marrying rather than just stating a proposition.

Beyond these prototypical cases, the literature has quickly broadened the understanding of the notion of SA as a special type of linguistic object that encompasses questions, orders and assertions and transcends propositional content revealing communicative intentions on the part of the speaker (Bach and Harnish, 1979; Gunlogson, 2008; Asher and Lascarides, 2008; Giannakidou and Mari, 2021). Speech acts can in-

deed be understood as *attitudes* towards propositional content: by *asserting* the speaker presents the propositional content as true, by *questioning* the speaker reveals uncertainty towards propositional content, by *ordering* the propositional content is asked to be realized and with *exclamatives*, the speaker reveals some type of subjective evaluation towards propositional content. SA have received an extensive body of work in the computational linguistics literature (Stolcke et al., 2000; Keizer et al., 2002; Carvalho and Cohen, 2005; Joty and Mohiuddin, 2018) and have shown to be an important step in many down stream NLP applications such as dialogues summarization (Goo and Chen, 2018) and conversational systems (Higashinaka et al., 2014). In this paper we attempt to measure for the first time the role of SA on urgency detection in tweets, focusing on natural disasters.

Previous works on communicative intentions during emergency crises has focused on the correlation between specific topics and SA (Zhang et al., 2011; Vosoughi, 2015; Elmadany et al., 2018a; Saha et al., 2020). As discussed below, it has been observed that people globally tend to react to natural disasters with SA distinct from those used in other contexts (e.g. celebrities): we might hypothesize that this is because all SA are not suited to constitute a pertinent reaction to emergency. Here, we explore the further hypothesis that SA can moreover be used as a sorting key between urgent and non-urgent utterances made in the same context of reaction to a natural disaster.

Before moving to real scenarios that rely on SA-aware automatic detection of urgency (this is left for future work), we first want to (1) measure the impact of SA in detecting intentions during crisis events in manually annotated data, and (2) explore the feasibility of SA automatic detection in crisis corpora. To this end, we present here an annotation schema for tweets using speech acts that (i) takes into account the variety of linguistic means whereby speech acts are expressed (including lexical items, punctuation, etc), both at the message and sub-message level, (ii) newly uses a two-level classification of speech acts, and (iii) intersects a classification of urgency where messages are annotated according to three classes (non useful vs. urgent vs. non urgent). Our contributions are:

- A new annotation scheme of speech acts in tweets at two levels of granularity (message and sub-message levels) that goes beyond flat classification of SA used in related work.
- A new French dataset of 6,669 tweets annotated for both urgency and SA, extending a first layer of urgency annotations initially proposed by Kozlowski et al. (2020).[3] [4]
- A qualitative and quantitative analysis of the annotation campaign, highlighting the correlation between SA and urgency categories.
- A set of deep learning experiments to detect SA in social media content relying on transformer architectures coupled with relevant linguistic features about how SA are linguistically expressed.

This paper is organized as follows. Section 2 presents related work in SA detection in social media as well as main existing crisis datasets. Section 3 provides the classification of SA we propose and the annotation guidelines to annotate them. Section 4 details the dataset we relied on and the results of the annotation campaign. Section 5 focuses on the experiments we carried out to detect SA automatically. We end by some perspectives for future work.

2. Related Work

2.1. Crisis Datasets

The literature on emergencies detection has been growing fast in the recent years and several datasets (mainly tweets) have been proposed to account for crisis related phenomena.[5] Messages are annotated according to relevant categories that are deemed to fit the information needs of various stakeholders like humanitarian organizations, local police and firefighters. Well known dimensions include relatedness (also known as usefulness or informativeness) to identify whether the message content is useful (Jensen, 2012), situation awareness (also known as urgency, criticality or priority) to filter out on-topic relevant (e.g., immediate post-impact help) vs. on-topic irrelevant information (e.g. supports and solicitations for donations) (Imran et al., 2013; McCreadie et al., 2019; Sarioglu Kayi et al., 2020; Kozlowski et al., 2020), and eyewitnesses types to identify direct and indirect eyewitnesses (Zahra et al., 2020). Annotations in most existing datasets are usually done at the text level. Some studies propose to additionally annotate images within the tweets (see for example (Alam et al., 2018)).

The question of how speakers convey emergency at the sentence level, has nonetheless be only tangentially addressed in a literature that has considered the correlation between specific speech acts and specific topics, without overtly addressing what the speech act shape of urgent messages is (see below).

2.2. Speech Acts in Social Media

Some amount of attention has been indeed devoted to understanding how speech acts (as used on Twitter) vary qualitatively according to the topic discussed or *topic*. In this line of questioning, SA have been studied as filters for new topics.

[3] https://github.com/DiegoKoz/french_ecological_crisis

[4] If accepted, the dataset will be publicly available.

[5] See https://crisisnlp.qcri.org/ for an overview.

Zhang et al. (2011) in particular, resorts to a Searlian typology of SA that distinguishes between assertive **statements** (description of the world) and expressive **comments** (expression of a mental state of the speaker). Zhang et al. (2011) also distinguish between interrogative **questions** and imperative **suggestions**. Finally, a category **miscellaneous** brings together the Searlian **declaratives** and the **commissives**, used to make promises. Concerning the question of emergency, Zhang et al. (2011) showed that the SA's distribution on Twitter in the context of a natural disaster (e.g. earthquake in Japan) is distinctive: it is essentially composed by statements, associated to comments and suggestions / orders. In this context new information or ideas on how to (re)act are indeed expected and assertions are the most suitable to this aim. By contrast, discussion over a celebrity will mostly generate comments and almost no order or suggestion. Indeed, in this context, subjectivity matters more than immediate action.

Also inspired by Searle's typology, Vosoughi (2015; Vosoughi and Roy (2016) distinguish six categories: **Assertions, Recommendations, Expressions, Question Requests** and **Miscellaneous**. The authors use the definitions of Zhang et al. (2011), by distinguishing the *topic* discussed in the tweets, from the *type* of topic (*Entity-oriented*, *Event-oriented topics*, or *Longstanding topics*). 6 topics were then selected (2 of each type): for *entity-oriented*, they are interested in Ashton Kusher and the Red Sox; for *event-oriented*, they study the Boston bombings in 2013 and the Ferguson demonstrations in 2014; for *Longstanding topics*, they consider cooking and travel. The distribution of speech acts that the authors obtain allows them to show that there is a greater similarity of distribution between topics of the same type than between topics of different types. On the other hand, the *entity-oriented* and *event-oriented* types are closer to each other, with a majority of assertions and expressions, whereas for the *longstanding* types, assertions are less abundant and recommendations well represented.

In this same perspective of topic identification, Elmadany et al. (2018b) classify 21,000 tweets in Arabic according to their topic type and distinguish events (for example, in our case, natural disasters), entities (especially people) and various issues such as travel or cooking. Each tweet is associated to a pair of speech act/sentiment according to the following classification: **Assertions, Recommendations, Expressions and Requests**, and among Sentiments, the standard Positive, Negative, Mixed and Neutral categories. Their study makes emerge a salient association between assertions and people/events and neutrality on the one hand and an association between expressivity long-standing topics and negativity on the other.

For completeness, we note that SA have been studied in the context of political campaigns, notably by Subramanian et al. (2019) (The 2016 Australian "federal election cycle"), with a corpus composed of official statements / tweets / press clippings (Subramanian et al., 2019), where each statement is associated with a SA and a target party (liberal or conservative). The categorization envisioned by the authors articulates **Assertives**, **Commissives-action-specific**, **Commissive-action-vague**, **Commissives-outcome** (about a future reality state), **Directives**, **Expressives**, **Past-actions** and **Verdictives** (an assessment on prospective or retrospective actions). They observe an over-representation of assertives (40%), followed by verdictives (25%) and specific action (12%). The other categories represent less than 10% of the annotations. It is interesting to note that if we remove their precise characterization, commissives represent a little less than a quarter of the assigned speech acts, whereas they are almost absent from our corpus whose topic is emergency.

As far as we are aware, communicative intentions have never been explored in the context of urgency detection. This paper fills this gap by crossing the urgency classification and the SA classification in order to elucidate the interactions between speaker's attitudes and urgency categories (and their associated actions).

To achieve this, and as never previously undertaken, we propose a two-level typology of speech acts that allows us to characterize both the message as a whole and its parts providing multiple handles to study the correlation between emergency and speakers intentions.

3. Classification and annotations

We developed two sets of annotations: (i) one level classification including four distinct categories to label the tweet as an atomic unit, and (ii) a two-level annotation including the first four level categories and 8 second level categories. The second-level categories are used to annotate tweets at the subtweet level as opposed to the tweet as a whole. The goal of this double annotation both at the level of the tweet and at the level of the subtweet allows us to gather data to understand which part of the tweets determines the speech act at the holistic level.

3.1. Tweet level

Our classification of SA elaborates on the fundational Austinian and later Searlian distinction by (i) relying on propositional content and lexical clues such as modals (*should*, *must*, *can*, ...), evaluative adjectives, attitude verbs (*think*, *believe*, *want*, *hope* ...); (ii) introducing the category 'subjectives', which reshuffles some of the earlier classifications ('wishes', for instance are 'subjectives' rather than 'jussives' in our classification (e.g. (Condoravdi and Lauer, 2012)); (iii) considering presuppositional content as well (see (Mari, 2016) on French).

We distinguish four first-level categories which are mutually exclusive and define tweets as wholes, at a holistic level, as shown in Figure 1.

(1) JUSSIVES, as defined by (Zanuttini et al., 2012), enhance commitment to take action, as in (2)

(2) #Inondation Si vous êtes en zone inondable, découvrez comment préparer un kit de survie

(#Flooding If you are in an area at risk of flooding, discover how to prepare a survival kit).

In our classification we distinguish: *commissives* (i.e. the speaker commits himself or herself), *exhortatives* (i.e. the speaker commits some relevant individuals), *orders* (i.e. the speaker commits the addressee, in the case of authority relations), and *open-options* (i.e. the speaker describes the existence of a possibility).

(2) ASSERTIVES. Assertions are considered to convey objective truth (as opposed to subjective truth (Giannakidou and Mari, 2021)). With assertives, the speaker is committed toward the truthfulness of the proposition that is being uttered ((Portner, 2018) a.o.) and require their interlocutor to update the common ground (Ginzburg, 2012).

(3) Inondations dans l'Aude : la région débloque 25M€, le président Macron sur place lundi
(Flooding in Aude: the region unlocks 25M€, the president Macron on the spot on Monday).

(3) INTERROGATIVES. This category is dedicated to a variety of questions including both those that require an informative answer and those that, besides triggering an answer, reveal bias and expectations on the part of the speaker (see (Ladd, 1981)).

(4) Salut Chelsea, comment ça va, la tempête, par chez vous?
(Hi Chelsea, how is the storm at your place?).

(4) SUBJECTIVES. Finally, with subjectives, the speaker shares a mental state that can be either a personal evaluation or preference (see among many others (Lasersohn, 2005)) or an expressive state (an emotion or a feeling). The interlocutor is asked to update the common ground not just with the content of the evaluation but with the evaluation itself (see (Simons, 2007), and for recent discussion on French (Mari and Portner, 2021)). In our classification, 'wishes', for instance are 'subjectives' rather than 'jussives' as they do not trigger any committment to act so to make the content of the wish true.

(5) Grosse pensée à ma Laure qui est en Martinique avec l'ouragan
(My thoughts are with my Laure, who is in Martinique with the hurricane.)

Finally, **OTHERS** is added to the classification, for uncertain or unclassifiable cases.

(6) Simulation #3D d'une #inondation à Issy-les-Moulineaux merci à @Ubick3D pour le prêt #ortho3D #InterAtlas
(3D simulation of a flood in Issy-les-Moulineaux thanks to @Ubick3D for the loan #ortho3D #InterAtlas).

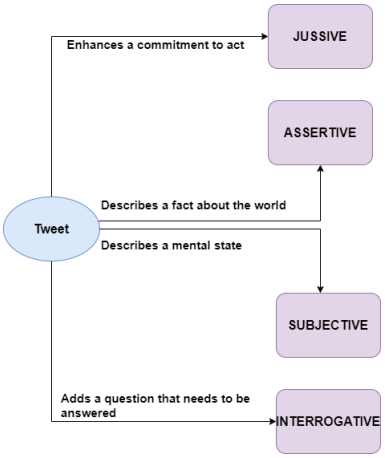

Figure 1: A classification for tweets that makes use of four illocutionary categories.

3.2. Segment level

The two-level annotation considers that each tweet is a small discourse composed of one or more statements, so that it can not only be classified at the holistic level but also at the level of its segments (we identify them between '[...]'). In order to achieve this, we have maintained the classification above at the holistic level, and we have elaborated each of the four categories to annotate the tweets at the segment level. At the segment level we use eight categories (see Figure 2), some of them are inspired by (Core et al., 1998) (e.g. the open-options).

From JUSSIVES, the annotation makes the distinction between (a) **OPEN-OPTIONS** – the speaker puts forward a possibility and leaves the addressee free to realize it or not (cf. (7))– , and (b) utterances that enhance a direct commitment on the part of a discourse participant, *ie*. **COMMISSIVES, EXHORTATIVES, ORDERS AND INTERDICTIONS**, that are called **OTHER-JUSSIVES** (cf. (8)).

(7) Ouragan #Irma : victime des intempéries ? **[OPEN-OPTION Conseils déclaration de sinistre par téléphone et en ligne @MAIF]**
(Hurricane #Irma : victim of the bad weather ? [OPEN-OPTION Claim reporting tips by phone and online @MAIF])

(8) Une grosse pensée pour les familles des victimes. **[OTHER-JUSSIVE Taxons le carbone dès maintenant pour éviter que les choses empirent dans le futur.]**
(A big thought for the families of the victims. [OTHER-JUSSIVE Let's tax the carbon now to prevent things from getting worse in the future.])

In ASSERTIONS, both second-level categories are deter-

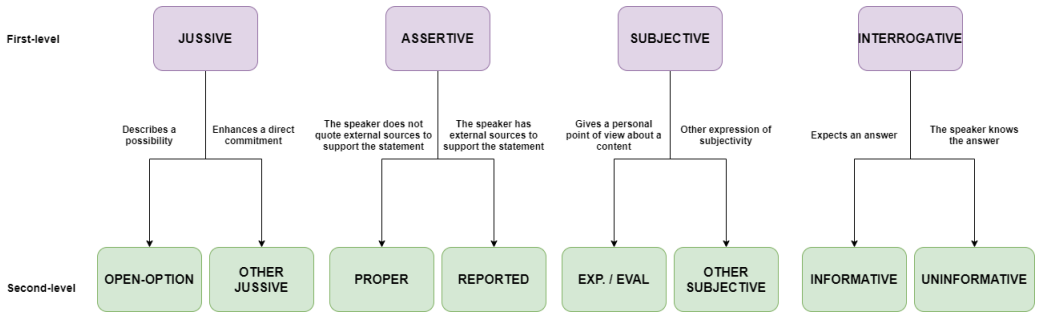

Figure 2: Two-layers annotation for tweets and inner segments.

mined by the source of knowledge that the speaker relies upon, i.e. the evidentiality condition as defined by (Saurí and Pustejovsky, 2009). If the speaker grounds their utterance on a third-party source, the assertive utterance is (a) a **REPORTED ASSERTION**, whereas if there is no such explicit source, it is a (b) **PROPER ASSERTION**, see (9) and (10) respectively.

(9) **[REPORTED Des patrouilles de police mises en place pour dissuader les cambrioleurs, via @franceinfo]**
([REPORTED Police patrols implemented in order to deter intruders, via @franceinfo])

(10) [**PROPER Au printemps, la fonte rapide de la neige peut provoquer une inondation.]** Votre famille est-elle prête?
([PROPER During spring, the rapid melting of snow can cause flooding.] Is your family ready?)

In SUBJECTIVES, the distinction was made between (a) **EXPRESSIVES/EVALUATIVES** whereby the speaker describes a personal evaluation or an expressive state that it is not deemed to become common ground or truth (cf. (11)) and (b) **OTHER SUBJECTIVE** for utterances that do not explicitly fall in the previous category (eg: puns, greetings...).

(11) **[EXP./EVAL. Pensées pour les saint Martinois et particulièrement pour ma famille installée la bas]**
([EXP./EVAL. Thoughts for the Saint Martinois and especially for my family living there])

(12) **[OTHER-SUBJECTIVE Bonjour de la Guadeloupe !]** Oui effectivement la situation ici est dramatique.
([OTHER-SUBJECTIVE Hello from Guadeloupe!] Yes indeed the situation here is dramatic.)

In INTERROGATIVES, the distinction was made between (a) **INFORMATIVE** questions to which the speaker cannot answer, and the ones that are (b) **UNINFORMATIVE** when the speaker is biased towards an answer, as in (13) and (14) respectively.

(13) @EmmanuelMacron Où sont les renforts censés arrivés à Saint-Martin et **[INFORMATIVE que comptez-vous faire]**
(@EmmanuelMacron Where are the reinforcements that are supposed to arrive in St. Martin and [INFORMATIVE what are you going to do])

(14) seisme ressenti en guadeloupe **[UNINFORMATIVE pouvez vous confirmer svp]**
(earthquake felt in guadeloupe [UNINFORMATIVE could you please confirm])

A single message can be annotated with several labels at the segment level, with each segment being annotated by only one tag, as shown in (15). Here an **INTERROGATIVE tweet** composed of two segments : **a PROPER assertion** followed by an **UNINFORMATIVE question**.

(15) [INTERR. [PROPER seisme ressenti en guadeloupe] [UNINF. pouvez vous confirmer svp]]
([INTERR. [PROPER earthquake felt in guadeloupe] [UNINF. could you please confirm]])

4. Data and Annotation

4.1. Dataset

Since our focus is on crises that occur in metropolitan France and its overseas departments, we rely on the only available corpus of French tweets by (Kozlowski et al., 2020)[6] composed of 12,826 tweets collected using dedicated keywords about ecological crises that occurred in France from 2016 to 2019 and posted 24h before, during (48h) and 72h after the crisis: 2 floods that occurred in Aude and Corsica regions, 10 storms (Béryl, Berguitta, Fionn, Eleanor, Bruno, Egon, Ulrika, Susanna, Fakir and Ana), 2 hurricanes (Irma and Harvey), and 1 sudden crisis (Marseille building collapse). The dataset comes with additional metadata including: number of likes and retweets of the tweet, and number of likes, followers, following of the user.

[6] https://github.com/DiegoKoz/french_ecological_crisis

In this dataset, each tweet is annotated according to an urgency classification composed of three categories: URGENT that applies to messages mentioning human/infrastructure damages as well as security instructions to limit these damages during crisis events, NOT URGENT that groups support messages to the victims, critics or any other messages that do not have an immediate impact on actionability but contribute in raising situational awareness, and finally NOT USEFUL for messages that are not related to the targeted crisis or information pertaining to events occurring outside the French territories.

The collection is extremely imbalanced with 1,442 (11.24%) useful but NOT URGENT, 2,147 (16.74%) URGENT and 9,237 (72.02%) NOT USEFUL messages, which is in line with the proportions reported in other crisis corpora (see Section 2.1).

4.2. Results of the Annotation Campaign

A subset of this dataset composed of $6,669$ tweets has been selected for SA annotations, so that almost all URGENT (2,080) and NON URGENT (1,401) messages have been annotated. Only 3,188 NOT USEFUL tweets have been selected in order to reduce the size of this class but keep it majoritary. Note that pre-existing urgency tags and metadata information have been removed, this will prevent annotators to get biased by specific urgency-SA pairs.

We hired two native French speaking annotators, both master's degree students in Linguistics in order to annotate the tweets. We performed a two-step annotation with an intermediate analysis of agreement and disagreement between the annotators. 448 tweets have been annotated in the first step by both annotators so that the inter-annotator agreement could be computed (Cohen's Kappa=0.62). Most cases of disagreement come from the difficulty of disentangling SUBJECTIVES from ASSERTIVES, in particular when attitudes and modal expressions are used such as *believe*, *think that*, etc. Indeed, both the subjective expressions (*think*, *believe*, or even more complex modal-tense-aspect combinations as *fallait* (which translates as 'should have been' with an additional implicature of preference in (16))) or its content can be targeted, according to their contextual relevance. This delicate distinction is often resolved in different manners by annotators.

(16) Et maintenant il n'y a presque plus de fumée... Il fallait arrêter le trafic ce matin et pas au milieu de la journée.
(And now there is almost no more smoke... Traffic should have been stopped this morning and not in the middle of the day).

The final distribution of annotated tweets is 59.8%, 22.3%, 10%, 4.5% and 3.3% for ASSERTIVE, SUBJECTIVE, JUSSIVE, OTHER and INTERROGATIVE respectively. Table 1 details the frequency of the first layer SA tags (i.e., tweet level) when paired with the original urgency annotations. Concerning the two most frequent SA (ASSERTIVE and SUBJECTIVE), two observations emerge: (1) Among URGENT messages (resp. NON URGENT), 86.6% (resp. 48.7%) are ASSERTIVE; and (2) Only 5% of URGENT messages are SUBJECTIVE while 29% of NON URGENT messages are. Similarly, we observe that 7% of JUSSIVE are URGENT vs. 14% NON URGENT. All these frequencies are statistically significant using the $\chi 2$ test ($\chi 2 = 1,1011.62,\ df = 8$, $p < 0.01$). When measuring the dependency strength between urgency and SA categories using the Cramer's V, we get ($V = .28,\ df = 2$) which confirms the statistical correlation between these two classifications.

	URG	NON URG	NON USEF	TOTAL
ASSERT.	1,802	682	1,506	**3,990**
JUSS.	145	203	321	**669**
SUBJ.	106	406	976	**1,488**
INTERR.	20	58	145	**223**
OTHER	7	52	240	**299**
Total	**2,080**	**1,401**	**3,188**	**6,669**

Table 1: Urgency- First layer SA annotation pairs statistics.

Table 2 further details the SA distribution for each crisis. We can see that ASSERTIVE messages are the most frequent ones regardless of the crisis. Another interesting finding concerns the distribution of SA in sudden crises. Indeed, SA frequencies are relatively similar in natural disaster crises (flood, storms and hurricane) with about 60% of ASSERTIVE and 20% of SUBJECTIVE. However in the Marseille building collapse, we observe a higher proportion of SUBJECTIVE (35% vs. 49% for ASSERTIVE) showing that people tend to express fewer messages of warning-advice but many critics denouncing the lack of effectiveness of government social action.

Table 3 presents the percentage of the second layer SA tags when paired with urgency labels[7] while Table 4 gives the distribution of the most frequent sequences showing that the most frequent single tags are also the most frequent when in first position of a sequence. This suggests that relying only on the first tag of a multi-label sequence can be relevant.

An in-depth analysis of the distribution of frequencies in these tables reveals that assertivity — and most prominently PROPER ASSERTIONS — is an indication of urgency. This means that speakers privilege (what they consider) truthful information over orders and commands to enhance action (on the part of the rescuing teams, for instance). Indeed, in our classification ASSERTIONS do not include subjective evaluations, and thus convey content informationally reliable

[7]Note that the frequencies of SA tags in this table are statistically significant ($\chi 2 = 710.70,\ df = 14, p < 0.01$).

		ASSERTIVE	SUBJECTIVE	JUSSIVE	INTERROGATIVE
Flood	Aude	718	184	84	20
	Autre	631	180	137	28
	Corse	248	73	45	23
	Total	1,597	437	266	71
Storms	Beryl	174	87	22	11
	Bruno	201	94	17	15
	Susanna	230	92	45	6
	Ulrika	170	60	43	7
	Berguitta	189	73	35	14
	Fionn Corse	238	69	28	6
	Egon	185	95	24	10
	Eleanor	208	69	26	7
	Total	1,595	639	240	76
Hurricane	Harvey	168	59	36	23
	Irma	487	251	100	36
	Total	655	310	136	59
Collapse	Marseille	143	102	27	17

Table 2: SA distribution for each crisis.

	URG	NON URG	NON USEF
JUSSIVE			
open-opt.	5.79	8.78	8.41
other.	7. 85	6.96	5.31
ASSERTIVE			
report.	**15.41**	7.84	7.81
proper.	**60.80**	39.63	45.01
INTERROGATIVE			
infor.	0.22	1.66	2.42
uninfor.	1.23	3.90	4.90
SUBJECTIVE			
eval/exp.	6.89	**28.36**	19.14
other.	1.80	2.85	7.00

Table 3: Urgency- Second layer SA annotation pairs in percentage.

and objectively veridical (i.e. conform to the outer reality and not a mental state) (Giannakidou and Mari, 2017; Giannakidou and Mari, 2018; Giannakidou and Mari, 2021) and thus ready for uptake and endorsement (e.g. (Ginzburg, 2012), (Krifka, 2019)) on the part of those who will bring help. The fact that speakers privilege PROPER ASSERTIONS to indicate urgency reveals that they are fully committed to the truthfulness of the message , of which they present themselves as the primary informational source. We note that assertions are also very frequent in non-useful messages. This is the case when the message contains information that is irrelevant to the crisis.

On the contrary, we can observe that SUBJECTIVES correlate with absence of urgency. Among subjectives EVALUATIVES/EXPRESSIVES are largely used to convey truths that are relativized to a 'judge' or an individual (a.o. (Lasersohn, 2005; Stephenson, 2007)) and are not eligible to function as reliable information for the rescuing services. A minority of subjectives encompass attitudes, whereby truth is also relativized to a particular mental state and cannot (without further negotiation) immediately become common ground (e.g. (Gunlogson, 2008; Mari and Portner, 2021)) and be ready for uptake on the part of the helpers.

Second-level SA	**%**	
PROPER ASSERTION	20.69	42.47
PROPER ASSERTION + other(s) SA	21.78	
EVAL./EXPR.	8.24	21.36
EVAL./EXPR. + other(s) SA	13.12	
REPORTED ASSERTION	5.22	7.68
REPORTED ASSERTION + other(s) SA	2.46	
OPEN-OPTION	5.04	6.69
OPEN-OPTION + OTHER(S) SA	1.65	

Table 4: Distribution of most frequent second-level SA.

5. Automatic Detection of SA

Now the dataset has been annotated, the next step is to automatically detect SA. We explore here the detection at the tweet level, leaving the second level for future work. Most state of the art approaches make use of feature-based machine learning algorithms (SVM, Naive Baise) relying on various surface, lexicon and syntactic features such as unigrams, punctuations, POS, emoticons and sentiment words (Vosoughi and Roy, 2016; Zhang et al., 2011; Algotiml et al., 2019). Deep learning architectures have also been explored, mainly for Arabic SA detection (Elmadany et al., 2018b) and English tweets relative to political campaigns (Subramanian et al., 2019) or topic oriented events (Saha et al., 2020). As far as we know, this is the first attempt to detection SA in a French crisis dataset. To this end, we train different classifiers on 80% of the instances of our dataset and test them on 20%. Note that the OTHER instances (around 300 tweets) have been removed from the dataset for the experiments as

they are very less frequent in urgent tweets and have no regular linguistic patterns. The final dataset is therefore composed of 6,370 tweets.
We experiment with several feature-based (SVM) and deep learning models (CNN, BiLSTM, transformers) but we only report here the models having the best results.

- **BERT$_{\text{base}}$** relies on the pre-trained BERT multilingual cased model (Devlin et al., 2019). We used the HuggingFace's PyTorch implementation of BERT (Wolf et al., 2019) that we trained for four epochs.
- **FlauBERT$_{\text{base}}$** and **CamemBERT$_{\text{base}}$** use respectively the FlauBERT (Le et al., 2019) and the CamemBERT base cased models (Martin et al., 2020), two pre-trained French contextual embeddings. We run them for four epochs and a learning rate of $2e-5$.
- **CamemBERT$_{\text{focal}}$** This model is similar to CamemBERT$_{\text{base}}$, but it uses focal loss (Lin et al., 2017) instead. Our aim here is to compare with one of the most effective approach for handling imbalanced data (Cui et al., 2019).
- **CamemBERT$_{\text{base}}$+F** and **CamemBERT$_{\text{focal}}$+F**. We finally experimented with multi-input models that use extra-features added on top of pre-trained contextual word embeddings, among which[8]: the presence of URLs, punctuation (exclamation marks and question marks) and the presence of numbers, as they are often used in tweets to indicates phone numbers of emergency rescue services or weather forecast.

Table 5 gives the results obtained in terms of precision (P), recall (R) and macro-F1 (F). Among the models we proposed, CamemBERT$_{\text{focal}}$+F that combine dedicated features with focal loss to handle class imbalance achieves the best with an F1 score of 73.55. When looking into the detailed results per class (cf. Table 6), we observe that ASSERTIVE and SUBJECTIVE classes are well predicted compared to JUSSIVE and INTERROGATIVE. An error analysis shows that despite having four classes, over half of our errors come from the ASSERTIVE class. We can explain that because the classifier will prioritize the most represented class, that is why ASSERTIVE scores over 80% in recall. Indeed, the objective-subjective distinction is often not clearcut and the ASSERTIVE will be preferred by the system.

(17) Laurent Dumonteil entouré de la Bretagne, on y est !!! (Lautent Dumonteil surrounded by brittany, here we are!!!)
Gold= SUBJECTIVE, Predicted= ASSERTIVE

[8]We also tested several other features including tweet meta-features, sentiment and emoticons, number of imperatives verbs, etc., but the results were not conclusive.

Finally, we observed that the classifier is misguided by the interrogation mark. Indeed, 60% of the errors where INTERROGATIVE has been wrongly predicted include tweets containing at least one interrogation mark, as in the example below.

(18) Comment un avion peut atterrir dans une tempête qui empêche les bagages de sortir ? C'est pas possible xD (How can a plane land in a storm that prevents the luggage from getting out? It is not possible xD)
Gold = SUBJECTIVE, Predicted= INTERROGATIVE

Models	P	R	F
BERT$_{\text{base}}$	64.81	58.00	60.80
FlauBERT$_{\text{base}}$	72.13	66.19	68.80
CamemBERT$_{\text{base}}$	74.16	70.57	71.22
CamemBERT$_{\text{base}}$+F	75.26	70.47	72.64
CamemBERT$_{\text{focal}}$	75.23	71.62	72.22
CamemBERT$_{\text{focal}}$+F	**75.66**	**71.95**	**73.55**

Table 5: Overall SA classification results.

	P	R	F
ASSERT.	87.06	88.72	87.89
JUSS.	75.22	60.28	64.44
SUBJ.	72.93	77.10	66.93
INTERR.	67.44	61.70	74.96
Accuracy=81.87			

Table 6: Best model results per class.

6. Conclusion

In this paper, we presented the first corpus-based study to measure the impact of speech acts in messages posted during crisis events in social media. We first proposed a new annotation guideline to annotate speech acts both at the tweet and subtweet levels, then a new dataset annotated for both speech acts and urgency categories in French. Our results show a strong correlation (i) between Assertive messages (in particular those that rely on first hand knowledge, i.e. PROPER ASSERTIONS) and urgency and (ii) Subjective messages and absence of urgency, with a high frequency of expressives and evaluatives. We finally conducted a set of experiments to detect SA at the tweet level relying on transformer architectures augmented with dedicated features. Our results are encouraging and constitute the first step towards SA-aware urgency detection in social media content.

Acknowledgment

This work has been supported by the INTACT project funded by the AAP CNRS - INHESJ 2022 and the FIESPI grant with the French Interior Ministry. Alda

Mari gratefully acknowledges ANR-17-EURE-0017 FrontCog. The research of Farah Benamara and Véronique Moriceau is also partially supported by DesCartes: the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) program.

7. Bibliographical References

Alam, F., Ofli, F., and Imran, M. (2018). CrisisMMD: Multimodal Twitter Datasets from Natural Disasters. *arXiv:1805.00713 [cs]*, may.

Algotiml, B., Elmadany, A., and Magdy, W. (2019). Arabic tweet-act: Speech act recognition for Arabic asynchronous conversations. In *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, pages 183–191, Florence, Italy, August. Association for Computational Linguistics.

Asher, N. and Lascarides, A. (2008). Commitments, beliefs and intentions in dialogue. In *Proceedings of the 12th Workshop on the Semantics and Pragmatics of Dialogue*, pages 29–36. Citeseer.

Austin, J. L. (1962). *How to do things with words*. Oxford University Press.

Bach, K. and Harnish, R. M. (1979). Linguistic communication and speech acts.

Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Harvard university press.

Carvalho, V. R. and Cohen, W. W. (2005). On the collective classification of email "speech acts". In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '05, page 345–352, New York, NY, USA. Association for Computing Machinery.

Condoravdi, C. and Lauer, S. (2012). Imperatives: Meaning and illocutionary force. *Empirical issues in syntax and semantics*, 9:37–58.

Core, M., Ishizaki, M., Moore, J., Nakatani, C., Reithinger, N., Traum, D., and Tutiya, S. (1998). The report of the third workshop of the discourse resource initiative, chiba university and kazusa academia hall. Technical report.

Cui, Y., Jia, M., Lin, T.-Y., Song, Y., and Belongie, S. (2019). Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Association for Computational Linguistics.

Elmadany, A., Mubarak, H., and Magdy, W. (2018a). Arsas: An arabic speech-act and sentiment corpus of tweets. In Hend Al-Khalifa, et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France, may. European Language Resources Association (ELRA).

Elmadany, A., Mubarak, H., and Magdy, W. (2018b). Arsas: An arabic speech-act and sentiment corpus of tweets. *OSACT*, 3:20.

Giannakidou, A. and Mari, A. (2017). Epistemic future and epistemic must: nonveridicality, evidence, and partial knowledge. In *Mood, Aspect, Modality Revisited*, pages 75–118. University of Chicago Press.

Giannakidou, A. and Mari, A. (2018). A unified analysis of the future as epistemic modality. *Natural Language & Linguistic Theory*, 36(1):85–129.

Giannakidou, A. and Mari, A. (2021). *Truth and veridicality in grammar and thought: Mood, modality, and propositional attitudes*. University of Chicago Press.

Ginzburg, J. (2012). *The interactive stance*. Oxford University Press.

Goo, C.-W. and Chen, Y.-N. (2018). Abstractive dialogue summarization with sentence-gated modeling optimized by dialogue acts. In *Proceedings of 7th IEEE Workshop on Spoken Language Technology*.

Gunlogson, C. (2008). A question of commitment. *Belgian Journal of Linguistics*, 22(1):101–136.

Hamblin, C. L. (1970). Fallacies. *Tijdschrift Voor Filosofie*, 33(1).

Higashinaka, R., Imamura, K., Meguro, T., Miyazaki, C., Kobayashi, N., Sugiyama, H., Hirano, T., Makino, T., and Matsuo, Y. (2014). Towards an open-domain conversational system fully based on natural language processing. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 928–939, Dublin, Ireland, August. Dublin City University and Association for Computational Linguistics.

Imran, M., Elbassuoni, S. M., Castillo, C., Diaz, F., and Meier, P. (2013). Extracting information nuggets from disaster-related messages in social media. *Proc. of ISCRAM, Baden-Baden, Germany*.

Jensen, G. E. (2012). Key criteria for information quality in the use of online social media for emergency management in New Zealand. Master's thesis.

Joty, S. and Mohiuddin, T. (2018). Modeling speech acts in asynchronous conversations: A neural-CRF approach. *Computational Linguistics*, 44(4):859–894, December.

Keizer, S., op den Akker, R., and Nijholt, A. (2002). Dialogue act recognition with Bayesian networks for Dutch dialogues. In *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue*, pages

88–94, Philadelphia, Pennsylvania, USA, July. Association for Computational Linguistics.

Kozlowski, D., Lannelongue, E., Saudemont, F., Benamara, F., Mari, A., Moriceau, V., and Boumadane, A. (2020). A three-level classification of french tweets in ecological crises. *Information Processing & Management*, 57(5):102284.

Krifka, M. (2019). Commitments and beyond. *Theoretical Linguistics*, 45(1-2):73–91.

Ladd, D. R. (1981). A first look at the semantics and pragmatics of negative questions and tag questions. In *Papers from the... Regional Meeting. Chicago Ling. Soc. Chicago, Ill*, number 17, pages 164–171.

Lasersohn, P. (2005). Context dependence, disagreement, and predicates of personal taste. *Linguistics and philosophy*, 28(6):643–686.

Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., Allauzen, A., Crabbé, B., Besacier, L., and Schwab, D. (2019). FlauBERT: Unsupervised Language Model Pre-training for French. *arXiv preprint arXiv:1912.05372*.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988.

Mari, A. and Portner, P. (2021). Mood variation with belief predicates: Modal comparison and the raisability of questions. *Glossa: a journal of general linguistics*, 40(1).

Mari, A. (2016). Assertability conditions of epistemic (and fictional) attitudes and mood variation. In *Semantics and Linguistic Theory*, volume 26, pages 61–81.

Martin, L., Muller, B., Ortiz Suárez, P. J., Dupont, Y., Romary, L., Villemonte de La Clergerie, É., Seddah, D., and Sagot, B. (2020). CamemBERT: a Tasty French Language Model. In *ACL 2020 - 58th Annual Meeting of the Association for Computational Linguistics*, Seattle / Virtual, United States, July.

McCreadie, R., Buntain, C., and Soboroff, I. (2019). TREC incident streams: Finding actionable information on social media. In Zeno Franco, et al., editors, *Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management, València, Spain, May 19-22, 2019*. ISCRAM Association.

Olteanu, A., Vieweg, S., and Castillo, C. (2015). What to Expect When the Unexpected Happens: Social Media Communications Across Crises. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pages 994–1009.

Portner, P. (2018). *Mood*. Oxford University Press.

Sadock, J. (2004). 3 speech acts. *The handbook of pragmatics*, page 53.

Saha, T., Patra, A. P., Saha, S., and Bhattacharyya, P. (2020). A transformer based approach for identification of tweet acts. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Sarioglu Kayi, E., Nan, L., Qu, B., Diab, M., and McKeown, K. (2020). Detecting urgency status of crisis tweets: A transfer learning approach for low resource languages. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4693–4703, Barcelona, Spain (Online), December. International Committee on Computational Linguistics.

Saurí, R. and Pustejovsky, J. (2009). Factbank: a corpus annotated with event factuality. *Language resources and evaluation*, 43(3):227.

Searle, J. R. (1975). Indirect speech acts. In *Speech acts*, pages 59–82. Brill.

Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua*, 117(6):1034–1056.

Stephenson, T. (2007). Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and philosophy*, 30(4):487–525.

Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Van Ess-Dykema, C., and Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3):339–374.

Subramanian, S., Cohn, T., and Baldwin, T. (2019). Target based speech act classification in political campaign text. *arXiv preprint arXiv:1905.07856*.

Vieweg, S., Castillo, C., and Imran, M. (2014). Integrating Social Media Communications into the Rapid Assessment of Sudden Onset Disasters. In *Proceedings of the 6th International Conference of Social Informatics*, SocInfo'14, pages 444–461.

Vosoughi, S. and Roy, D. (2016). Tweet acts: A speech act classifier for twitter. *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016)*.

Vosoughi, S. (2015). *Automatic detection and verification of rumors on Twitter*. Thesis, Massachusetts Institute of Technology.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., and Brew, J. (2019). HuggingFace's Transformers: State-of-the-art Natural Language Processing. *ArXiv*, abs/1910.03771.

Zahra, K., Imran, M., and Ostermann, F. O. (2020). Automatic identification of eyewitness messages on twitter during disasters. *Information Processing & Management*, 57(1):102–107.

Zanuttini, R., Pak, M., and Portner, P. (2012). A syntactic analysis of interpretive restrictions on imperative, promissive, and exhortative subjects. *Natural Language & Linguistic Theory*, 30(4):1231–1274.

Zhang, R., Gao, D., and Li, W. (2011). What are tweeters doing: Recognizing speech acts in twitter.

In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*.