Boróka Tímea Pápay,* Bálint György Kubik,**
Júlia Galántai*** & Károly Takács****

Intersections. EEJSP
8(4): 149–178.
https://doi.org/10.17356/ieejsp.v8i4.939
https://intersections.tk.hu

Gossip is distinct from other topics in spontaneous conversation

* [papay.boroka@tk.hu] (Doctoral School of Sociology, Corvinus University of Budapest)
** [kubikbalint@gmail.com] (Centre for Social Sciences, Budapest)
*** [galantai.julia@tk.hu] (Centre for Social Sciences, Budapest)
**** [takacs.karoly@tk.hu] (Linköping University; Centre for Social Sciences, Budapest)

Abstract

Gossip – talking about relevant others in their absence – is believed to constitute a large part of informal communication. The perception of the prevalence of gossip implies that it can be unambiguously identified and distinguished from other topics in spontaneous conversation. Its distinctiveness may be justified by multiple theoretical perspectives, including one that describes in-group gossip as an informal device for enforcing norms and punishing norm violators, and another that claims that gossip is used to release frustration and communicate envy. If the ultimate reason for gossip is to facilitate social bonding between the sender and the receiver, however, this would not differentiate gossip from other conversational topics that provide social enjoyment, such as entertainment and food. In a novel contribution, we explore the topics included in a corpus containing 550 hours of unfiltered spontaneous conversation and identify using LDA topic modeling whether some topics are unambiguously prominent in in-group gossip. The explorative approach is integrated with the manual annotation of instances of gossip across the entire corpus. We identified coherent topics of in-group gossip that are clearly different from those of small talk and storytelling. Our analysis finds that feelings, intentions, and opinions are frequently expressed in in-group gossip, more than habits, manners, and behavior. In-group gossip topics are characterized by more words associated with anger, in line with theoretical perspectives that attribute the motives of norm enhancement and punishment or frustration and envy to gossip.

Keywords: gossip; social bonding; maintenance of norms; spontaneous conversations; LDA topic model

1 Introduction

1.1 The prevalence of gossip in spontaneous conversation

Humans are empowered with exceptional and complex verbal language skills. Probably most important of all the associated benefits, spoken language helps us obtain information about events, wrong-doings, and expectations within our social group. Receiving information about others helps us relate properly to actions and behavior and to make the right decisions in a complex social world (e.g., Suls, 1977; Giardini, 2012).

Talking about others who are not present is called *gossip* (Foster, 2004; Kurland, 2011; Ellwardt, 2011; Beersma & van Kleef, 2012; Beersma et al., 2012; Grosser et al., 2012; Giardini & Wittek 2019a). Gossip is prevalent in schools (Kisfalusi et al., 2019; Estévez et al., 2022), organizations (Kurland & Pelled, 2000; Ellwardt, Labianca & Wittek, 2012; Beersma & van Kleef, 2012) and other contexts in life (Besnier, 2009; Mangardich et al., 2019; Martinescu et al., 2019; De Backer et al., 2019). According to some accounts, two-thirds of human verbal communication falls into this category (Emler, 1994; Dunbar, 2004; Dunbar et al., 1997).

The prevalence of gossip can only be confirmed if gossip is identifiable and can be differentiated from other informal speech acts. But is gossip different in its content and semantic features from other conversation topics? This is a fundamental question that research first needs to answer. If gossip is indeed different, what are the distinct characteristics of gossip-based conversations?

While the importance of gossip has been demonstrated in previous qualitative, experimental, and survey research, knowledge is limited about whether gossip is differentiable from other conversation topics and characterizable in terms of distinct characteristics and semantic features. As a novel contribution, we explored the topics in unfiltered spontaneous conversations in the HuTongue corpus we built for this purpose using a quantitative explorative approach to identify major topics of informal communication. We labeled the topics based on their characteristic words. We explored the topics that emerged and checked whether they could be characterized as in-group gossip topics and if they stood out unambiguously from others. Based on the theoretical literature that anticipates certain substantive properties of gossip, we quantitatively and qualitatively contrasted in-group gossip topics with other topics.

1.2 Theoretical explanations for and perspectives about the motivation for gossip

The expected distinctiveness of gossip conversations can be explained by various theoretical accounts. We review these accounts in this subsection by grouping them into perspectives emphasizing *social bonding*, *the maintenance of social norms*, *social undermining*, and *emotion venting*. Some of these theoretical accounts aim at providing ultimate explanations, while others offer proximate explanations and highlight the individual sources of motivation for gossip. These theoretical views imply complementary predictions about the distinctive characteristics of gossip conversations compared to informal interactions without gossip.

First, the distinctiveness of gossip conversations is not evident if ultimate explanations for gossip are considered. One widely shared perspective is that gossip, similarly to dance, music, and other rituals, facilitates *social bonding* within the group (Dunbar, 1993; 1997; 1998; 2004; 2021). In this sense, gossip is part of 'social grooming' that is used to release stress, create close contact, and make life enjoyable (Dunbar, 1998). The fundamental characteristics of gossip conversations, such as intimacy, close distance, confidentiality between the sender and the receiver, and a high level of enjoyment and excitement (Feinberg, Willer & Schultz, 2014) provide support for this theoretical perspective and underline its relevance in human evolution. Social bonding in gossip is also clearly reflected in the public image of gossip as idle talk. Adopting this perspective, talking about third parties has a similar purpose to

small talk about weather, food, entertainment, and sex (Levin & Arluke 1985). The targets of gossip and consequences for the latter are considered less important than strengthening the relationship between the sender and receiver from the perspective of social bonding.

Second, considering information gathering and validation, gossip might transmit evaluative information about others as an efficient alternative to direct observation (Bozoyan & Vogt, 2016). Gossip that transmits reputational information might be beneficial or detrimental to the target (Wu, Balliet & van Lange, 2015; 2016a; 2016b; 2016c). Reputational information exchange could be linked to group protection and cooperation, as it may punish norm violators and free riders in a subtle way (Fine, 1977; Feinberg et al., 2012; 2014; Nowak & Sigmund, 2005; Giardini & Vilone, 2016; Jazaieri et al., 2019; Giardini & Wittek, 2019c; Számadó et al., 2021; Giardini et al., 2022). This theoretical perspective considers gossip to be an informal mechanism used to maintain group norms by decreasing the reputation of norm violators (Hess, 2006; Sommerfeld et al., 2008; Beersma & van Kleef, 2011; Giardini & Conte, 2012; Giardini et al., 2014; Feinberg et al., 2014; Giardini & Wittek, 2019b). From this perspective, the importance of gossip lies in its target-sanctioning potential. Third, due to its efficiency at altering reputations, 'senders' could be motivated to use gossip as a means of negative influence for their own benefit. For the sender, gossip could represent a form of purposeful action that harms the reputation of the target, who has a conflict of interest with the sender (Galen & Underwood, 1997; Paquette & Underwood, 1999; Crick et al., 1996). Such gossip behavior could be considered a form of social undermining that is designed to spoil the reputation of the target (Duffy et al., 2002; 2012; Dijkstra et al., 2014; Jeuken et al., 2015; Crick et al., 2001; Faris, 2012; Ellwardt, Labianca & Wittek, 2012). Fourth, gossip could be driven by emotional motives. At the individual level, such motives could be linked to releasing stress, broadcasting emotions (Harber & Cohen, 2005; Harber et al., 2014), or coping with envy. Experimental research shows that confidential gossip discussions are often used to liberate the sender from emotional burdens and have physiological consequences such as normalizing the pulse rate or causing excitement for the recipient (Beersma & van Kleef, 2012; Feinberg et al., 2014). *Emotion venting* – the desire to share emotionally evocative experiences – has been found to be a relevant motive for gossip that was previously overlooked in survey research (Pauw et al., 2018; Dores Cruz et al., 2019). Organizational studies have shown that negative emotions such as anxiety, disappointment, anger, and depression are frequent consequences of negative gossip that can hinder the achievement of organizational goals as they decrease job satisfaction and increase employee fluctuation (Hobfoll, 1989; Agnew, 1992), and may also be responsible for the decay of cooperation (Giardini & Wittek, 2019c). In short, there are various reasons to support the claim that gossip is not different from small talk. Other theoretical accounts, however, highlight that in-group gossip may be clearly distinguishable from other conversation topics such as storytelling, talking about food, entertainment, or politics. The fundamental question of the distinctiveness of gossip gains importance from the prevalence of gossip in human life and should be put to the empirical test.

1.3 Research methods for studying gossip

Research on gossip has benefited from multiple methodologies. While experimental methods are straightforward to use to investigate the relationship of gossip to cooperation (Feinberg et al., 2014; Samu et al., 2020; Samu & Takács, 2021) and to reveal its physiological correlates

(Beersma & van Kleef, 2012; Feinberg, 2012), survey research can tell a lot about the perceptions of and motivation for gossip (Ellwardt, Labianca & Wittek, 2012; Lyons & Hughes, 2015; Kisfalusi et al., 2019). Research has also analyzed the presence of gossip in interviews, social media, workplace emails, and surveys (Mitra, 2012). The abstract situational context of laboratory experiments and reluctance to provide information on confidential gossip in survey research limit the external validity of research with these methods. Research on gossip in spontaneous conversations is very much needed to answer questions about its true nature. Studies on gossip in natural settings are based on anthropological observations (Gluckman, 1963; Levin & Arluke, 1985; Besnier, 2009; Emler, 1994; Dunbar et al., 1997), but there have been a few attempts to take account of gossip in transcribed conversations (Slade, 1997; Foster, 2004; Eckhaus & Ben-Hador, 2019; Robbins & Karan, 2020; Szabó et al., 2021).

The analysis of actual conversations helps determine whether gossip stands out from other conversation topics and may be characterized as having distinct characteristics and semantic features. In this study, we quantitatively explore the topics of unfiltered spontaneous conversations in the HuTongue corpus that we built for this purpose. We devote particular attention to phrases concerning entertainment, food, and other small talk that could be related to social bonding.

This explorative approach provides valid results about the distinctiveness of gossip, as we use it in combination with the costly manual annotation of gossip instances. The distinctiveness of topics that are labeled in-group gossip in the explorative strategy can be validated if segments of gossip topics strongly correlate with the instances of in-group gossip identified by manual annotation. Once such validation occurs, the distinctive characteristics of in-group gossip topics can be analyzed. In this way, our study tests the claim of the distinctiveness of gossip and contributes to understanding the prevalence of gossip in informal human communication.

1.4 Expectations about the characteristics of gossip topics

Although gossip may be defined in several ways (Dores Cruz et al., 2020), gossip with and about in-group members may be particularly important in specific theoretical accounts. Gossip with in-group members is expected to increase social bonding and identification with the group. Gossip about in-group members is essential for social orientation in the group about everyday encounters and informs members about the violation of group norms. In our manual annotation, we focused on in-group gossip defined as conversation between at least two group members about a third group member who was not present during the conversation. This might or might not have an evaluative element (cf. Dores Cruz et al., 2020) and excludes *out-group gossip,* which is about targets who are not members of the group, and *storytelling,* which is about past events that happened *to the speakers themselves.*

Table 1 summarizes theoretical expectations about the distinctive characteristics of gossip topics broken down by the theoretical accounts and explanations outlined in Subsection 1.2.

Table 1 Theoretical perspectives on gossip and related characteristics

social bonding, social enjoyment	correlated with small talk topics, entertainment, love, food, weather, and politics; funny storytelling
maintenance of norms and punishment of norm violators, group protection, reputational information exchange	Discussion of manners, personal habits, role performance, task interdependence, deviance, and other reputational concerns; exemplary storytelling
negative influence, undermining	bullying, moral downgrading, offense, exclusion
emotion venting, release of frustration, envy	self-defense, pain, desires, misfortune

What kind of characteristics can be expected of gossip conversations that are in line with the various theoretical accounts and motivations? Social enjoyment of gossip (first row in Table 1) would imply similarities with topics that individuals enjoy discussing, such as entertainment, love, food, weather, and politics. Storytelling could also be of this character as it is often regarded as a process that encourages talking and listening, and as a trigger for starting and continuing conversations with the aim of reflecting on experiences (Bruner, 1986; Labonté & Feather, 1996; Cheshire, 2000). Storytelling typically entertains the receiver and in exchange improves the attractiveness and the perceived status of the sender (Donahue & Green, 2016; Bietti, Tilston & Bangerter, 2018). Storytelling could be a socially collaborative activity through which people focus on sharing personal memories (Mandelbaum, 2013; Bietti et al., 2018; Bietti et al., 2019) and could be related to the concepts of episodic future thinking and episodic memory, which are based on an individual's ability to remember past personal experiences and then recall them as events that might happen in the future (Tulving, 2002; Schacter, 2017). Storytelling, however, as a description of a past series of events or future or hypothetical scenarios with moral or emotional implications, could also create and propagate group norms. Its non-routine character guides behavior in uncertain and novel situations and strengthens group norms and identity (Labov & Fanshel, 1977; Coates, 1996; Cheshire, 2000; Husnu, Mertan & Cicek, 2018; Bietti, Tilston & Bangerter, 2018). Hence, the motivation for storytelling might have similarities in terms of a desire for the maintenance of group norms (second row). As gossip transmits reputational information about other individuals, it is expected to contribute to the establishment and maintenance of social order (Beersma & van Kleef, 2011; Giardini & Conte, 2012; Ellwardt, Labianca & Wittek, 2012; Giardini et al., 2014; Giardini & Wittek, 2019b; Feinberg et al., 2014; Hess, 2006). Hence, the emerging topics of in-group gossip are most likely to cover personal habits, manners, appearance, and the role performance of the target (Levin & Arluke, 1985; Giardini & Wittek, 2019b). From an analysis of spontaneous speech, Levin & Arluke (1985) concluded that the topics and subjects of gossip are mainly personal habits, manners, appearance, and role performance. While other parts of conversation may directly prescribe obedience to group norms and their facilitation, talking about other group members could posit role models to follow and behaviors to avoid. As the reputation of the target is directly altered through gossip, the normative content of gossip is expected to be especially relevant if the target is an in-group member. If the target is not part of the social context (e.g., a film star or a family member), gossip could still serve the function of prescribing norms, but it is less likely to alter the in-group reputational structure.

Gossip could also be driven by the individual-level motive of undermining the reputation of others (Duffy et al., 2002; Dijkstra et al., 2014). *Negative influence* (third row in Table 1) could potentially be achieved by belittling the target and questioning their goodwill, relationships, intentions, and behavior. A target can be depicted to be incompetent, immoral, or evil, be attributed a questionable action, or explicitly insulted. As a result, the speakers might question the legitimacy of the group membership of the target (Wert & Salovey, 2004).

Emotional expressions and non-verbal forms of emotional communication could also occur in gossip for other reasons (fourth row in Table 1). *Emotion venting*, the desire to share emotionally evocative experiences, is one important motive for gossip (Pauw et al., 2018; Dores Cruz et al., 2019). Emotion venting may be traced in the use of language associated with strong emotions. Text analysis has been successful at social emotion detection by focusing on the sentiment associated with individual words, as these play a central role in how we describe and understand emotions (Shimanoff, 1985; Shaikh, Prendinger & Ishizuka, 2008; Kazemzadeh, Lee & Narayanan, 2013; Correa, Scherman & Arriagada, 2016). It has been shown that negative emotions are more likely to occur in indirect forms of speech (Anderson, 1998). Relying on established text analysis strategies, we use both an emotion dictionary and the annotation of non-verbal emotional expressions to examine the distinctiveness of in-group gossip topics according to these dimensions.

In line with this motivation, we explore whether the topics of in-group gossip can be characterized by different emotional, sentimental, and procedural features than other topics. In our analytical strategy, we first explored the topics of informal conversation in our unfiltered speech corpus and tested if gossip topics are clearly different from non-gossip topics. Finding that they are, we examined if the substantive features of gossip topics differ quantitatively from other topics.

2 Data and methods

2.1 Context and data

The data we use is from a Hungarian TV reality show. Participants during this time were restricted within a closed environment and had almost no possibility to interact with the outside world. Participants were competing for a final prize and left the context after one-on-one duels. Some activities were organized, but all conversations occurred naturally, and no conversations were pre-scripted. Contracts between the entertainment company and fully consenting volunteer participants included detailed information about the presence of full-time audio recording.

Uninterrupted high-quality audio recordings were made using the personal microphones of participants. Video recordings were unavailable. Only edited summaries of 30 minutes duration of daily events were broadcast on television. The entertainment company provided us with the audio data for scientific research with a non-disclosure agreement. We analyzed data covering a period of eight consecutive days following the middle of the 105-day competition. Fifteen participants started the competition, and eight were still participating during this period. We selected this period because the competition had not yet entered its final phase but participants had spent sufficient time together to get to know each other well, so their conversations could be considered natural.

We manually transcribed and annotated all conversations and built a corpus from approximately 550 hours of unfiltered spontaneous conversations (HuTongue). We partitioned the corpus into segments separated by silences lasting longer than two seconds (Galántai et al., 2018).

Although the context of the investigation implies the occurrence of certain topics such as competing and the selection of competitors, the topics and characteristics of conversations closely resembled those found in everyday talk, such as discussions about family, friends, and intimate relations. However, because of the specific context, conversations may have been more competitive and involved more acting up for outside viewers than is natural. We should note, however, that natural conversations also involve a large proportion of acting for audiences (Goffman, 1978).

Table 2 Annotations tag used for the corpus

Emotional expressions	*Interpretative codes*	*Presence of third-person codes*
coughing **(k)** sighing **(s)** laughing **(~)** crying **(*)** mocking laughter **(gn)** confused laughter **(zn)** yawning **(a)** hiss **(pi)** screaming **(sik)** elation **(u)** throat clearing **(tor)** whistling **(f)** singing **(é)**	Incomprehensible word **(?)** Unclear word **(( ))** Presence of distant speakers, but their speech is incomprehensible **(t?)** Presence in the conversation is unidentifiable. An estimate of the number of participants present is given, for example: **(4)** Speaker is not a participant **(k?)** To mark the end of sound effects and interpreting codes: **(code)))**	Speakers talking about a third person **(p)** Speakers talking about a certain person who is a participant, for example about Sean: **(p-S)** Someone present in the conversation, but remains silent or taking part in another conversation, for example, Sean **(S)** Someone talking among the participants, for example: **(Sean)**

2.2 Annotation

We used a complex manual transcription and annotation strategy using the software f4 (Dr. dresing & pehl GmbH, Marburg, Germany, https://www.audiotranskription.de/english/f4). Annotation took place during transcription. Annotators used time stamps to indicate the exact time interval of speech events and documented which participant(s) were talking and for how long. Name tags provided information about turn-taking and simultaneous speaking situations. Annotators always marked the names of speakers in the conversation (e.g., Sean). Names were anonymized after transcription. Annotators were extensively trained to use tags. The quality of the transcription and annotation was tested and ensured in several ways (for details, see Supplementary File S1).

Several non-verbal expressions were annotated (Table 2), including lowered voice, laughter, crying, sighing, coughing, and throat clearing. We asked annotators to indicate incomprehensible, unidentifiable speech, and uncertainty about the recipients of the speech act.

2.3 Tagging gossip

Annotators coded conversations about third persons who were *perceived* as not being present as gossip (right-hand column in Table 2: 'p'). Third persons were perceived as absent if they did not take part in the conversation as speakers, were not addressed during the conversa-

tion by the other speakers, and their voice was not audible in the background. The perceived presence of participants who remained silent during the conversation was also annotated (e.g., S). In-group targets of gossip were also annotated (e.g., p-S). In-group members were defined as participants and former participants of the reality show.

Annotators were asked to use gossip tags for lines of the conversation when they perceived that speakers were talking about a person who was not present. Statements about the third person's deeds, personality, and numerous other factors all fall into this category. Annotators were also instructed to use the gossip tag if the speaker made a statement about him- or herself in relation to a third person who was perceived to be absent. When the speakers mentioned multiple participants who were not present in the conversation, all of them were marked individually as gossip targets.

Conversations in which the target was not present but was not a participant or former participant (such as acquaintances, relatives known only to the sender, or celebrities) were interpreted as out-group gossip for which the target was not tagged. Table 3 provides segment examples of in-group gossip and other informal talk in the corpus. As some examples illustrate, targets are not always mentioned by their names, but with pronouns that only manual annotation could interpret properly. In the last in-group gossip example, the sender (Grace) is talking about her relation to the target (Miranda) in a way that has evaluative content.

Table 3 Differentiation of in-group gossip in the annotation of the corpus (examples)

In-group gossip	**Not in-group gossip**
#23:18# (**Andrew**) She cooks only once or twice a week, but after that there is such a mess that we have to wash up after her for two weeks. (**p-G**) #23:21#	#22:48# (**Victoria**) Mum cooks quite well. She cannot bake though. #22:50#
#12:44# (**Sean**) I am disappointed with Kyle. I think he has lied to us several times. He even lied to me while looking me in the eyes. He often equivocates and looks for excuses. (**p-K**) #12:47#	#09:19# (**Grace**) When my brother was younger, he always wanted mum to make him breakfast. When I prepared it for him, he did not eat it. #09:21#
#14:58# (**Grace**) Miranda told me that you revealed to her you act strategically with not getting into conflicts or quarrels. I believe her, I don't think she lied to me. (p-M) #15:02#	#11:22# (**Victoria**) Put on some more weight, that's what mum said. She said that it looks good on me, as a woman needs to be in shape. #11:26#

2.4 Text pre-processing

The analysis of texts written in agglutinative languages like Hungarian requires the lemmatization of the corpus due to the large variety of potential suffixes for the same stems. We used magyarlanc ('Hungarian chain') to pre-process and lemmatize the HuTongue corpus (Zsibrita, Vincze & Farkas, 2013). magyarlanc is a tool for linguistic analysis that was developed for the syntactic analysis of Hungarian. We implemented part of the speech tagging with magyarlanc and received morphosyntactic information about words in the corpus as an output.

We developed a 'stopword' dictionary based on the translation of the stopword dictionary from the Snowball project (Porter, 2001). Using magyarlanc for morphological analysis and part-of-speech tagging, multiple stopwords were added to the dictionary. We excluded all words from our corpus that were not categorized into a known morphological category by magyarlanc (represented by 'X' in the program output). Adverbs, apart from verbal adverbs, were also discarded along with adpositions, auxiliary verbs, interjections, particles, determiners, and coordinating and subordinating conjunctions. A manual check by researchers ensured that foreign words used as normal parts of language and slang words incorrectly categorized by magyarlanc were not discarded unnecessarily. We also added other nonsensical words during manual qualitative checks. Our final stopword dictionary contained more than 2000 lemmas.

2.5 LDA topic modeling

Topic models are ideal for analyzing large unstructured collections of text (Blei, 2012; Mohr & Bogdanov, 2013; Colleoni, Rozza & Arvidsson, 2014; Németh, Katona & Kmetty, 2020). Topic models are often used in combination with other statistical tools to estimate differences between documents (Dimaggio, Nag & Blei, 2013) and identify patterns of language usage as they can highlight what people talk about (McFarland et. al., 2013). Topic modeling is also used for the content analysis of textual data to discover hidden themes based on word co-occurrence (Hagen, 2018) – for instance, in social media (Koltai, Kmetty & Bozsonyi, 2021; Vancsó & Kmetty, 2021) or large volumes of legislative text (Quinn et al., 2010). Topic modeling can detect and measure differences in the concentration of themes in a corpus. To identify underlying patterns, the model assigns observed words to topics and for each topic allocates high probability to few words from the given vocabulary (Dimaggio, Nag & Blei, 2013).

Topic model outputs require human judgment for interpretation (Mimno et al., 2011), and topics are often manually assigned with labels (Hall et al., 2008). With topic modeling, one can classify issues that occur during speech and combine these with key semantic features to describe conversation patterns and behavior during gossip (Bak, Lin & Oh, 2014).

Latent Dirichlet Allocation (LDA) (Blei, Ng & Jordan, 2003; Blei & Lafferty, 2006; Levy & Franklin, 2014) is a well-known and frequently used topic model. LDA is a generative statistical model that treats documents as a mixture of topics that are multinomial distributions of words (Hong & Davison, 2010; Vogel et al., 2012). The LDA topic model identifies sets of words that tend to reflect hidden topics that characterize every segment in the corpus (Blei, Ng & Jordan, 2003; Blei & Lafferty, 2006). This is based upon a 'bag-of-words' approach, which handles individual words as interchangeable. LDA models the term-topic and topic-document probabilities in a generative way with a Dirichlet distribution as a prior, then estimates non-exclusive topic memberships for each document in the corpus (Blei, Ng & Jordan, 2003; Blei & Lafferty, 2006).

Our pre-processed data of lemmatized text without stopwords were used as input for LDA topic modeling. Even though text pre-processing left a relatively small number of unique lemmas, terms appearing in less than five documents and words present in more than 60 per cent of texts were removed to discard overly rare and overly frequent lemmas. Numbers were excluded. The final document-term matrix included 12,961 documents and 8,530 terms.

Gensim version 3.2.0, a topic modeling library for Python, and its connecting Visdom backend were used to construct the document-term matrix for LDA modeling (Rehurek & Sojka, 2010). We randomly split our corpus into train (50 per cent), test (25 per cent), and validation sets (25 per cent). Our models were configured to use an asymmetric prior learned from the data and to make 40 passes through the training data.

During the process of model building and choosing the number of topics, we relied on multiple metrics such as logarithmic perplexity (measured on the test and validation sets), Jaccard distances, and Kullback-Leibler differences between consecutive training steps, as well as the semantic coherence metric by Mimno et al. (2011). After consulting the literature and undertaking qualitative and quantitative assessments of our corpus, we decided to use 50 topics. Figure 1 displays the metrics that support our choice.

Afterward, we used a sentiment and an emotion dictionary (Szabó, 2014; Szabó & Morvay, 2015; Szabó & Vincze, 2015; Szabó, Vincze & Morvay, 2016) (see the description of dictionaries in Appendix B) to obtain basic correlations with tags, such as the gossip annotation tag in our corpus and topics obtained as a result of LDA topic modeling.

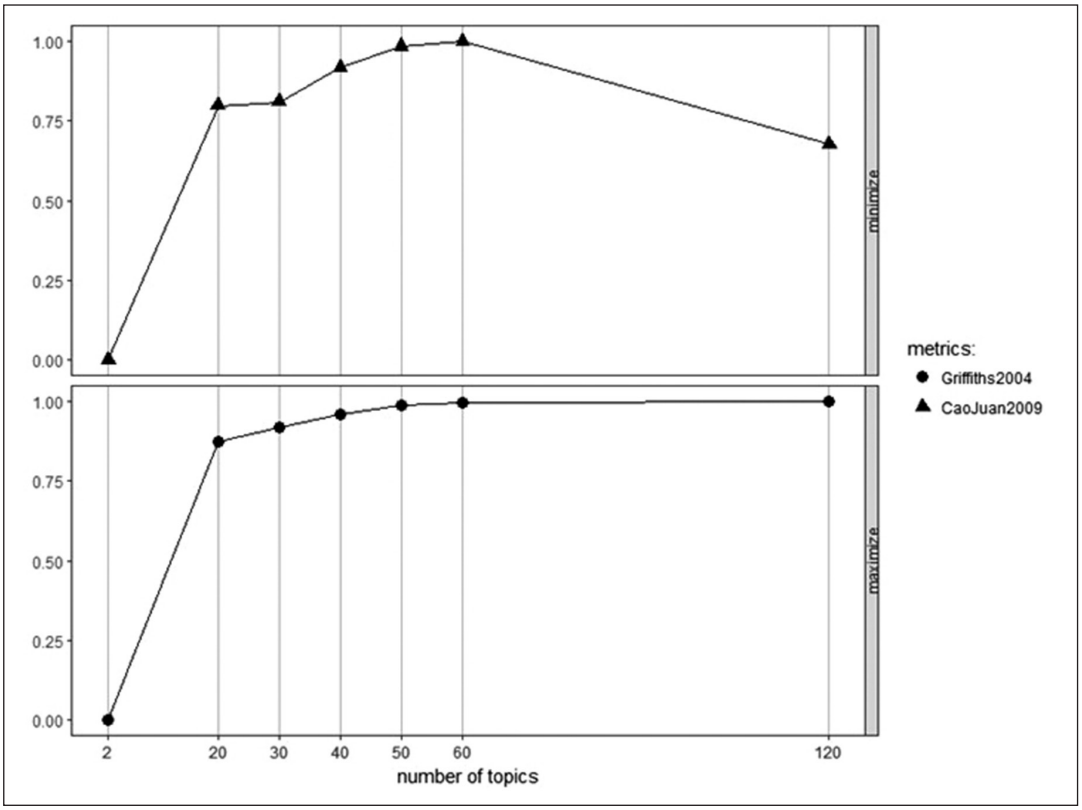

Figure 1 Index values for finding the optimal number of topics for the LDA model based on a density-based method (Cao et al., 2009) and a log-likelihood-based method (Griffiths & Steyvers, 2004).

Note: The density-based method (upper curve) maximizes the similarity in the intra-cluster and minimizes the similarity between clusters for adaptive LDA model selection. For the log-likelihood-based method (lower curve), the likelihood of the observed data is maximized by changing the number of topics. The log-likelihood is estimated with harmonic means using the Gibbs sampler.

Table 4 Topic summary and coherence by topic id. Average coherence is: -3.31.

Topic id	Topic coherence	Category	Topic id	Topic coherence	Category
0	−6.00	i	25	−4.55	i
1	−1.87	i	26	−3.38	s
2	−1.99	i	27	−3.13	i
3	−8.13	e	28	−2.74	i
4	−5.73	i	29	−10.08	i
5	−3.04	e	30	−1.36	i
6	−1.91	i	31	−1.60	?
7	−1.65	i	32	−1.59	g
8	−3.50	s	33	−1.63	i/e
9	−2.34	?/e	34	−1.96	e
10	−3.92	s	35	−1.80	e
11	−1.74	i	36	−2.56	i/e
12	−6.10	s	37	−6.31	s
13	−4.94	s	38	−2.40	e
14	−2.50	s	39	−2.54	i
15	−2.69	i	40	−5.97	s
16	−2.44	i	41	−1.21	g
17	−3.08	i	42	−2.18	s
18	−2.42	i	43	−1.08	i
19	−6.28	e	44	−1.76	s
20	−3.83	s	45	−1.59	i
21	−3.13	i	46	−3.50	e
22	−3.33	i	47	−8.58	s
23	−2.71	s	48	−1.90	e
24	−3.28	e	49	−1.54	s

Notes: Ex-post categorization: i = internal issues; e = entertainment; s = storytelling; g = in-group gossip.

3 Results

Table 4 summarizes the coherence of the 50 topics, and their categorization based on term weights into four main themes. Labels were assigned to the topics collaboratively by two different evaluators (Vogel et al. 2012). The rate of agreement of their labeling was 86 per cent. Consensus labeling identified four major themes of topics of spontaneous conversation: a) everyday life and 'internal issues' such as cooking and hygiene (N=21-23); b) topics about the reality show itself, including organized tasks and activities, the selection process, and duels (N=9-12); c) storytelling topics about the sender or a non-participant third person who was not present (N=14); and d), in-group gossip topics (N=2). Four topics were mixed or not coherent.

The two in-group gossip topics were distinctively about (the performance of) other group members and were most coherent. A desire for the maintenance of norms and punishment of norm violators could be traced in these topics, but feelings, intentions, and opinions appeared to be more prevalent than habits, manners, deviance, actions, and behavior. This indicates that perspective-taking is an important characteristic of in-group gossip (cf. Davis, 1983; Beersma et al., 2018; Righi & Takács, 2022).

Table 5 illustrates words characteristic of one topic about everyday issues, one storytelling topic, and two in-group gossip topics. Words associated with food are well represented in the first topic, while the second is filled with terms referencing outside parties (such as family members and celebrities), their activities, and associated feelings (e.g., love). Looking at the linguistic features of the topics, internal issues and topics related to the reality show equally involve verbs, nouns, and adjectives. Storytelling and out-group gossip topics contain more non-participant names (family, friends, and celebrities), personal pronouns, action verbs, adjectives, and nouns. In contrast, the two in-group gossip topics contain the names of participants, personal pronouns, and words that describe perceptions of others (such as feel, think, and understand) with large weights. There is evidence that participants often mention themselves while gossiping. Perception words often refer to the sender; mainly about their feelings or thoughts. The two in-group gossip topics contain nouns with lower weights.

We correlated the four main themes of the topics with some characteristics such that we could statistically associate these findings with the different theoretical expectations outlined in Table 1. The emergence of in-group gossip topics that are distinct from storytelling and out-group topics indicates that social enjoyment is not a distinctive characteristic of in-group gossip. Out-group gossip topics did not emerge independently of storytelling. They contained a total of twelve topics, characterized by larger weights for family names, celebrities, and public characters.

Table 5: Examples of topics and the most important words affiliated with them with their LDA weights

internal issues (topic 0)		***storytelling (topic 20)***		***in-group gossip (topic 32)***		***in-group gossip (topic 41)***	
little	0.04	*mom*	0.127	*tell*	0.065	*say*	0.083
know	0.034	*say*	0.039	*Kyle*	0.057	*you*	0.047
appetite	0.031	*mommy*	0.025	*know*	0.045	*him/her**	0.044
meat	0.029	*they*	0.023	*you*	0.042	*know*	0.030
egg	0.026	*bald*	0.020	*him/her**	0.038	*think*	0.022
ham	0.022	*bull*	0.017	*go*	0.036	*thing*	0.021
say	0.021	*real*	0.016	*Tommy*	0.028	*Zach*	0.018
depend	0.020	*show*	0.016	*Andrew*	0.024	*tell*	0.017
Diana	0.018	*small*	0.015	*Miranda*	0.023	*man*	0.014
bacon	0.017	*girl*	0.015	*compete*	0.018	*Miranda*	0.014
salt	0.017	*open*	0.014	*duel*	0.015	*opinion*	0.012
throw	0.016	*door*	0.013	*feel*	0.015	*understand*	0.011
Daniel	0.015	*know*	0.012	*sign*	0.014	*want*	0.011
thank	0.014	*eight*	0.012	*carry*	0.012	*feel*	0.010
sure	0.013	*trace*	0.011	*take*	0.011	*speak*	0.009
dress up	0.013	*doctor*	0.011	*Zach*	0.010	*Grace*	0.009
smooth	0.013	*brother/ sister**	0.011	*understand*	0.009	*stand*	0.009
let	0.012	*August*	0.011	*Daniel*	0.008	*keep*	0.009
issue	0.011	*ahh*	0.011	*put*	0.008	*love*	0.009
bored	0.011	*send*	0.010	*Ella*	0.008	*Sean*	0.008
love	0.011	*sure*	0.009	*pack*	0.008	*much*	0.007
draw	0.010	*neighbor*	0.009	*want*	0.008	*true*	0.007
independent	0.010	*do*	0.009	*Victoria*	0.007	*bad*	0.007
*him/her**	0.010	*hand*	0.009	*call*	0.007	*Victoria*	0.007
wait	0.010	*blond*	0.009	*game*	0.007	*Daniel*	0.006
fourth	0.010	*pick*	0.008	*do*	0.007	*nobody*	0.006
glue	0.010	*which*	0.008	*joke*	0.007	*honest*	0.006
happy	0.010	*born*	0.008	*stand*	0.007	*word*	0.006
salty	0.010	*shit*	0.008	*thing*	0.006	*wait*	0.006
dense	0.010	*nah*	0.008	*two*	0.006	*does*	0.006

Note: Words here are translated from Hungarian to English (words are lemmatized in Hungarian).

* these words have no masculine/feminine versions in Hungarian.

Table 6 Correlation coefficients (*r*) for topics and segment characteristics (segments are coherent units of conversation without silences longer than two seconds).

Segment characteristics	**internal (topic 0)**	**storytelling (topic 44)**	**in-group gossip (topic 32)**	**in-group gossip (topic 41)**
gossip ratio	−0.064 (p<0.001)	0.027 (p<0.001)	0.202 (p<0.001)	0.235 (p<0.001)
ratio of 'joyful' words	−0.015 (p=0.011)	−0.036 (p<0.001)	−0.058 (p<0.001)	−0.091 (p<0.001)
ratio of words associated with sadness	0.015 (p=0.014)	0.009 (p=0.134)	−0.010 (p=0.090)	−0.037 (p<0.001)
ratio of words associated with anger	0.007 (p=0.234)	−0.021 (p<0.001)	0.018 (p=0.003)	0.024 (p<0.001)
positive ratio	−0.023 (p<0.001)	−0.045 (p<0.001)	−0.067 (p<0.001)	−0.092 (p<0.001)
negative ratio	−0.020 (p<0.001)	−0.051 (p<0.001)	−0.058 (p<0.001)	−0.072 (p<0.001)
ratio of non-verbal annotation tags	−0.026 (p<0.001)	0.014 (p=0.018)	−0.046 (p<0.001)	−0.105 (p<0.001)
ratio of laughter annotation tags	−0.042 (p<0.001)	0.001 (p=0.915)	−0.073 (p<0.001)	−0.123 (p<0.001)
ratio of crying annotation tags	0.033 (p<0.001)	0.078 (p<0.001)	0.004 (p=0.558)	−0.048 (p<0.001)

Note: Four illustrative topics that emerged as a result of an LDA topic model based on a document-term matrix with 12,961 documents (segments) and 8,530 terms, df=12959.

We identified potential connections between different properties of individual segments and topics (Table 6). The first row of Table 6 confirms that in-group gossip topics contained text with more gossip tags. In-group gossip topics cannot be characterized by an abundance of words associated with social enjoyment; their appearance in fact is negatively correlated with in-group gossip topics. In-group gossip topics are characterized by a smaller proportion of laughter, a smaller proportion of sadness-related expressions, and fewer positive as well as negative sentiments than other topics. This indicates that in-group gossip topics are not distinct from other topics according to social enjoyment motivation. Furthermore, in-group gossip topics can be characterized by a greater frequency of words associated with anger. This is consistent with the theoretical perspectives that in-group gossip targets deviance and norm violations or aims at negative influence or is driven by emotion venting, envy, or frustration.

We analyzed all topics correlated with the variables such as the number of words in each topic (see Table 9 in the Appendix) and the number of persons present in conversations. Gossip topics seemed to involve fewer people. This maintains the idea that gossip is a confidential activity.

Appendix Table 9 displays bivariate correlations between selected topics from each theme and various quantitative segment characteristics, including the number of words, frequencies of annotations marks, and prevalence of words from sentiment and emotion dictionaries. Results in Appendix Table 9 indicate that in-group gossip segments contained significantly more words than others. The higher number of words per segment could be a sign of excitement (Rosnow, 2001). This result is also a consequence of the fact that conversations with in-group gossip are lengthier than others. Other correlations with the sentiment and emotion dictionaries and certain annotation marks indicate the relevance of emotion venting in in-group gossip and the release of frustration and envy in particular (last row in Table 1). Non-verbal emotional expressions that are indicators of social enjoyment such as laughing, sighing, and crying are associated with storytelling topics. This might indicate that storytelling and non-verbal emotions during communication are a kind of stage performance (Goffman, 1978).

4 Discussion

Gossip is a widespread activity that has been explained with reference to different sources of motivation and functions (Beersma & van Kleef, 2012; Farley, 2019; Dores Cruz et al., 2019; Emler, 2019). In this paper, we categorized different theoretical perspectives that define the function of gossip as *social bonding*, *maintenance of social norms*, *social undermining*, and *emotion venting*. In a very broad sense, and as an ultimate explanation, the human tendency to talk so much about others who are not present is due to the inclination for social bonding that makes our life enjoyable (Dunbar, 1998; 2004). Accordingly, social enjoyment has been identified as a key motivation for gossip in survey research (Beersma & van Kleef, 2012; Dores Cruz et al., 2019), but social enjoyment could also be a source of motivation for other conversations, such as about entertainment or food. Some proximate explanations of gossip and theoretical perspectives on its sources of motivation suggest that in-group gossip topics are distinct from topics of social enjoyment. In-group gossip could be used to exchange information about the reputations of others and to protect a group by helping spot norm violations and free riding (e.g., Feinberg et al., 2012; 2014; Giardini et al., 2014). Gossip may also be motivated by negative influence and a desire for undermining (Duffy et al., 2012) or could be used for emotion venting (Pauw et al., 2018; Dores Cruz et al., 2019), or to release frustration and communicate envy (cf. Liu et al., 2016). These different sources of motivation for gossip are not necessarily in competition (Ellwardt, Steglich & Wittek, 2012).

While the motivation for gossip may be revealed through survey research (Beersma & van Kleef, 2012; Dores Cruz et al., 2019), the place and the distinctiveness of gossip can only be explored through the analysis of spontaneous conversations. This is difficult, as large corpora of unfiltered natural conversations are rarely available, and there are none we are aware of that have been subject to the labor-intensive manual annotation of in-group gossip. Hence, the extraction of topics from a large natural language corpus of unfiltered communication and the specification of topics that feature in-group gossip could be considered a novel contribution of our study.

As we aimed at creating an objective view of in-group gossip, we did not rely on event samples, filtered discussions, or partial observations. Our data is from a closed environment in which all interactions were recorded for a relatively long period of time. Our goal was to

provide an overview of the dominant topics discussed by the participants, to identify relationships among topics, and to uncover associations between topics and important characteristics that involve gossip. We identified two topics unambiguously as in-group gossip topics. We validated this labeling by correlating the manual annotation of gossip with the topics that emerged. We found the strongest presence of in-group gossip tags in the two gossip topics that have emerged. Hence, we could clearly identify and differentiate in-group gossip from other topics in terms of its content and main features (cf. Goldsmith & Baxter, 1996), and we linked these with the theoretical perspectives that we categorized. In-group gossip topics were distinct from topics of storytelling and out-group gossip that might contain informal evaluative communication about a third party who was not known to the receiver.

In-group gossip topics were the most coherent among all the topics and contained more words, which are interesting findings. This indicates that in-group gossip topics involved the greatest similarity of words within the topic, implying a large degree of stability in the semantic content of gossip, which is achieved by exploiting a colorful vocabulary. Gossip topics involved beliefs and opinions about others surprisingly often and were about relationships with the target. Gossip topics contained more names, personal pronouns, and verbs related to personal perceptions of feelings and thoughts, while adjectives and nouns were less frequent. In contrast, storytelling topics contained non-participant names and plenty of function verbs.

In addition, we analyzed quantitative relationships between in-group gossip topics, annotation marks, and emotions identified with emotion dictionaries. We found that gossip not only involves informing people or setting norms in a group, but it may personally impact the speaker by unleashing anger and distress. Among other emotions, sadness and joy were more typical of storytelling rather than gossip. These results imply that social bonding is probably not the most important motivation for in-group gossip. The results suggest that it is storytelling and out-group gossip that may be more closely related to social bonding in conversations (Dunbar, 1998; 2004).

However, it is important to highlight the contextual limitations of our study. TV reality shows might not offer ideal conditions for supporting the evolutionary accounts of gossip, which may be better identified in small-scale societies (Besnier, 2009). Furthermore, the HuTongue corpus is unique in character as it covers an unprecedented amount of manually transcribed and annotated unfiltered spontaneous conversation, but in a very specific context involving only eight participants. The limitation of our corpus is that the conversations were influenced by a competition which took place during the period of recording. The latter theme probably occurred more frequently during the conversation than in everyday life situations, which may have influenced some of the topics that were identified during our analysis. The strongly competitive nature of the situation, however, does not necessarily mean that frustration and anger were greater because of this circumstance. To arrive at more general conclusions, our findings need to be validated in other collections of conversations in different milieus in the future.

A major value of the study is the construction and analysis of the large corpus of unfiltered spontaneous conversations. Furthermore, we have generated new insight for understanding gossip in everyday human conversations, but similar large-scale studies are needed to confirm our findings in other contexts, with more speakers, and in other languages. Subsequent analyses could identify the structure, the content, and the context of topics as well as their correlation with participants' presence and further relevant variables such as the time of day and activities carried out in parallel with speech.

References

Agnew, R. (1992). Foundation for a general strain theory of crime and delinquency. *Criminology, 30*(1), 47–88. https://doi.org/10.1111/j.1745-9125.1992.tb01093.x

Anderson, K. J. & Leaper, C. (1998). Emotion Talk Between Same-and Mixed-Gender Friends. Form and Function. *Journal of Language and Social Psychology, 17*(4), 419–448. https://doi.org/10.1177/0261927X980174001

Bak, J. Y., Lin, C. Y. & Oh, A. (2014). Self-disclosure topic model for classifying and analyzing Twitter conversations. In B. Pang & W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (pp. 1986–1996). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1213

Beersma, B., & van Kleef, G. A. (2011). How the grapevine keeps you in line: Gossip increases contributions to the group. *Social Psychological and Personality Science, 2*(6), 642–649. https://doi.org/10.1177/1948550611405073

Beersma, B. & van Kleef, G. A. (2012). Why people gossip: An empirical analysis of social motives, antecedents, and consequences. *Journal of Applied Social Psychology, 42*(11), 2640–2670. https://doi.org/10.1111/j.1559-1816.2012.00956.x

Beersma, B., Bechtoldt, M. N. & Schouten, M. E. (2018). When Ignorance Is Bliss: Exploring Perspective Taking, Negative State Affect and Performance. *Small Group Research, 49*(5), 576–599. https://doi.org/10.1177/1046496418775829

Besnier, N. (2009). Gossip and the Everyday Production of Politics. *University of Hawaii Press.*

Bietti, L. M., Tilston, O. & Bangerter, A. (2018). Storytelling as adaptive collective sensemaking. *Topics in Cognitive Science, 11*(4), 710–732. https://doi.org/10.1111/tops.12358

Bietti, L. M., Bangerter, A., Knutsen, D., Mayor, E. (2019) Cultural transmission in a food preparation task: The role of interactivity, innovation and storytelling. *PLoS ONE 14*(9), e0221278. https://doi.org/10.1371/journal.pone.0221278

Blei, D. M. (2012). Probabilistic Topic Models. *Communications of the ACM, 55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M. & Lafferty, J. D. (2006). Dynamic Topic Models. *Proceedings of the 23rd International Conference on Machine Learning* (pp. 113–120). Association for Computing Machinery. https://doi.org/10.1145/1143844.1143859

Blei, D. M., Ng, A. Y. J. & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3,* 993–1022.

Bozoyan, C. & Vogt, S. (2016). The impact of third-party information on trust: Valence, source, and reliability. *PLoS ONE, 11*(2), e0149542. https://doi.org/10.1371/journal.pone.0149542

Bruner, J. (1986). *Actual Minds, Possible Worlds.* Harvard University Press.

Cao J., Xia, T., Li, J., Zhang Y. & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing, 72*(7–9), 1775–1781. https://doi.org/10.1016/j.neucom.2008.06.011

Cheshire, J. (2000). The telling or the tale? Narratives and gender in adolescent friendship network. *Journal of Sociolinguistics, 4*(2), 234–262. https://doi.org/10.1111/1467-9481.00113

Coates, J. (1996). *Women Talk*. Blackwell.

Colleoni, E., Rozza, A. & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, *64*(2), 317–332. https://doi.org/10.1111/jcom.12084

Correa, T., Scherman, A. & Arriagada, A. (2016). Audiences and disasters: Analyses of media diaries before and after an earthquake and a massive fire. *Journal of Communication*, *66*(4), 519–541. https://doi.org/10.1111/jcom.12245

Crick, N. R., Bigbee, M. A. & Howes, C. (1996). Gender differences in children's normative beliefs about aggression: How do I hurt thee? Let me count the ways. *Child Development*, *67*(3), 1003–1014. https://doi.org/10.2307/1131876

Crick, N. R., Nelson, D. A., Morales, J. R., Cullerton-Sen, C., Casas, J. F. & Hickman, S. E. (2001). Relational Victimization in Childhood and Adolescence. I Hurt You through the Grapevine. In J. Juvonen & S. Graham (Eds.), *Peer harassment in school: The plight of the vulnerable and victimized* (pp. 196–214). The Guilford Press.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

De Backer, C. J. S., Van den Bulck, H., Fisher, M. L. & Ouvrein, G. (2019). Gossip and Reputation in the Media: How Celebrities Emerge and Evolve by Means of Mass-Mediated Gossip. In F. Giardini & R. P. M. Wittek (Eds.), *The Oxford Handbook of Gossip and Reputation* (pp. 324–342). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190494087.013.17

DiFonzo, N. & Bordia, P. (2007). Rumor, gossip and urban legends. *Diogenes, 54*(1), 19–35. https://doi.org/10.1177/0392192107073433

Dijkstra, M., Beersma, B. & van Leeuwen, J. (2014). Gossiping as a response to conflict with the boss: Alternative conflict management behavior? *International Journal of Conflict Management*, *25*(4), 431–454. https://doi.org/10.1108/IJCMA-08-2014-0059

DiMaggio, P., Nag, M. & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding, *Poetics*, *41*(6), 570–606. https://doi.org/10.1016/j.poetic.2013.08.004

Donahue, J. K. & Green, M. C. (2016). A good story: Men's storytelling ability affects their attractiveness and perceived status. *Personal Relationships*, *23*(2), 199–213. https://doi.org/10.1111/pere.12120

Dores Cruz, T. D., Balliet, D., Sleebos, E., Beersma, B., Van Kleef, G. A. & Gallucci, M. (2019). Getting a grip on the grapevine: Extension and factor structure of the motives to gossip questionnaire. *Frontiers in Psychology, 10,* 1190. https://doi.org/10.3389/fpsyg.2019.01190

Dores Cruz, T. D., Nieper, A. S., Testori, M., Martinescu, E. & Beersma, B. (2020). An Integrative Definition and Framework to Study Gossip. *Group & Organization Management, 46*(2), 252–285. https://doi.org/10.1177/1059601121992887

Duffy, M. K., Ganster, D. C. & Pagon, M. (2002). Social undermining in the workplace. *Academy of Management Journal*, *45*(2), 331–351. https://doi.org/10.2307/3069350

Duffy, M. K., Scott, K. L., Shaw, J. D., Tepper, B. J. & Aquino, K. (2012). A social context model of envy and social undermining. *Academy of Management Journal*, *55*(3), 643–666. https://doi.org/10.5465/amj.2009.0804

Dunbar, R. I. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, *16*(4), 681–694. https://doi.org/10.1111/j.1439-0310.1998.tb00103.x

Dunbar, R. I. (1998). *Grooming, gossip, and the evolution of language*. Harvard University Press.

Dunbar, R. I. (2004). Gossip in evolutionary perspective. *Review of General Psychology, 8*(2), 100–110. https://doi.org/10.1037/1089-2680.8.2.100

Dunbar, R. I. (2021). *Friends: Understanding the Power of our Most Important Relationships*. Little, Brown.

Dunbar, R. I., Marriott, A. & Duncan, N. D. (1997). Human conversational behavior. *Human Nature*, *8*(3), 231–246. https://doi.org/10.1007/BF02912493

Eckhaus, E. & Ben-Hador, B. (2019). Gossip and gender differences: a content analysis approach. *Journal of Gender Studies*, *28*(1), 97–108. https://doi.org/10.1080/09589236.2017.1411789

Ellwardt, L. (2011). *Gossip in organizations. A social network study*. PhD thesis, University of Groningen. https://pure.rug.nl/ws/portalfiles/portal/2551611/15complete.pdf

Ellwardt, L., Labianca, G. J. & Wittek, R.P.M. (2012). Who are the objects of positive and negative gossip at work? A social network perspective on workplace gossip. *Social Networks, 34*(2), 193–205. https://doi.org/10.1016/j.socnet.2011.11.003

Ellwardt, L., Steglich, C. & Wittek, R.P.M. (2012). The Co-evolution of Gossip and Friendship in Workplace. *Social Networks*, *34*(4), 623–633. https://doi.org/10.1016/j.socnet.2012.07.002

Emler, N. (1994). Gossip, reputation, and social adaptation. In R. F. Goodman & A. Ben-Ze'ev (Eds.), *Good Gossip* (pp. 117–138). University Press of Kansas.

Emler, N. (2019). Human Sociality and Psychological Foundations. In Giardini, F. & Wittek, R. P. M. (Eds.), *The Oxford Handbook of Gossip and Reputation* (pp. 47–68). Oxford University Press.

Estévez, J.L., Kisfalusi, D. & Takács, K. (2022). More than One's Negative Ties: The Role of Friends' Antipathies in High School Gossip. *Social Networks, 70,* 77–89. https://doi.org/10.1016/j.socnet.2021.11.009

Faris, R. (2012). Aggression, Exclusivity, and Status Attainment in Interpersonal Networks. *Social Forces, 90*(4), 1207–1235. https://doi.org/10.2307/41683164

Farley, S. (2019). On the Nature of Gossip, Reputation, and Power Inequality. In Giardini, F. & Wittek, R. P. M. (Eds.), *The Oxford Handbook of Gossip and Reputation* (pp. 343–358). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190494087.013.18

Feinberg, M. (2012). *Exploring the Virtues of Gossip: The Prosocial Motivations and Functions of Reputational Information Sharing*. PhD thesis, UC Berkeley. https://escholarship.org/uc/item/33t2f9js

Feinberg, M., Cheng, J. & Willer, R. (2012). Gossip as an Effective and Low-Cost Form of Punishment. *Behavioral and Brain Sciences*, *35*(1), 25. https://doi.org/10.1017/S0140525X11001233

Feinberg, M., Willer, R., & Schultz, M. (2014). Gossip and ostracism promote cooperation in groups. *Psychological Science, 25*(3), 656–664. https://doi.org/10.1177/0956797613510184

Fine, G. A. (1977). Social components of children's gossip. *Journal of Communication, 27*(1), 181–185. https://doi.org/10.1111/j.1460-2466.1977.tb01815.x

Foster, E. K. (2004). Research on Gossip: Taxonomy, Methods, and Future Directions. *Review of General Psychology, 8*(2), 78–99. https://doi.org/10.1037/1089-2680.8.2.78

Galántai, J., Pápay, B., Kubik, B., Szabó, M. K. & Takács K. (2018). A pletyka a társas rend szolgálatában. Az informális kommunikáció struktúrájának mélyebb megértéséért a Computational Social Science eszközeivel [Gossip in Service for Social Order – Using the Tools of Computational Social Science for a Deeper Understanding of the Structure of Informal Communication]. *Magyar Tudomány, 179*(7), 964–976. https://doi.org/10.1556/2065.179.2018.7.5

Galen, B. R. & Underwood, M. K. (1997). A developmental investigation of social aggression among children. *Developmental Psychology, 33*(4), 589–600. https://doi.org/10.1037/0012-1649.33.4.589

Giardini, F. (2012). Deterrence and transmission as mechanisms ensuring reliability of gossip. *Cognitive Processing*, *13*(2), 465–475. https://doi.org/10.1007/s10339-011-0421-0

Giardini, F. & Conte, R. (2012). Gossip for social control in natural and artificial societies. *Simulation*, *88*(1), 18–32. https://doi.org/10.1177/0037549711406912

Giardini, F. & Vilone, D. (2016). Evolution of gossip-based indirect reciprocity on a bipartite network. *Scientific Reports*, *6*, 37931. https://doi.org/10.1038/srep37931

Giardini, F. & Wittek, R. P. M. (Eds.) (2019a). *The Oxford Handbook of Gossip and Reputation*. Oxford University Press. https://10.1093/oxfordhb/9780190494087.001.0001

Giardini, F. & Wittek, R. P. M. (2019b). Silence is Golden. Six Reasons Why Gossip Does Not Spread. *Frontiers in Psychology*, *10*, 1120. https://doi.org/10.3389/fpsyg.2019.01120

Giardini, F. & Wittek, R. P. M. (2019c). Gossip, Reputation, and Sustainable Cooperation: Sociological Foundations. In F. Giardini & R. P. M. Wittek (Eds.), *The Oxford Handbook of Gossip and Reputation* (pp. 22–46). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190494087.013.2

Giardini, F., Paolucci, M., Villatoro, D. & Conte, R. (2014). Punishment and gossip: Sustaining cooperation in a public goods game. In B. Kamiński & G. Koloch (Eds.), *Advances in Social Simulation* (pp. 107–118). Springer. https://doi.org/10.1007/978-3-642-39829-2_10

Giardini, F., Balliet, D., Power, E. A., Számadó, Sz. & Takács, K. (2022). Four Puzzles of Reputation-Based Cooperation: Content, Process, Honesty, and Structure. *Human Nature, 33,* 43–61. https://doi.org/10.1007/s12110-021-09419-3

Gluckman, M. (1963). Gossip and scandal. *Current Anthropology*, *4*(3), 307–316. https://doi.org/10.1086/200378

Goffman, E. (1978). *The presentation of everyday life*. Harmondsworth.

Goldsmith, D. J. & Baxter, L. A. (1996). Constituting Relationships in Talk: A Taxonomy of Speech Events in Social and Personal Relationships. *Human Communication Research, 23*(1), 87–114. https://doi.org/10.1111/j.1468-2958.1996.tb00388.x

Griffiths, T. L. & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences, 101*(suppl. 1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Grosser, T. J., Lopez-Kidwell, V., Labianca, G. & Ellwardt, L. (2012). Hearing it through the grapevine: Positive and negative workplace gossip. *Organizational Dynamics*, *41*(1), 52–61. https://doi.org/10.1016/j.orgdyn.2011.12.007

Hagen. L. (2018). Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models? *Information Processing and Management, 54*(6), 1292–1307. https://doi.org/10.1016/j.ipm.2018.05.006

Hall, D., Jurafsky, D. & Manning, C. (2008). Studying the history of ideas using topic models. In M. Laputa & H. T. Ng (Eds.), *Proceedings of the 2008 Empirical Methods in Natural Language Processing* (pp. 363–371). Association for Computational Linguistics. https://aclanthology.org/D08-1038

Harber, K. D. & Cohen, D. J. (2005). The emotional broadcaster theory of social sharing. *Journal of Language and Social Psychology, 24*(4), 382–400. https://doi.org/10.1177/0261927X05281426

Harber, K. D., Podolski, P. & Dyer, L. (2014). Hearing stories that violate expectations leads to emotional broadcasting. *Journal of Language and Social Psychology, 33*(1), 5–28. https://doi.org/10.1177/0261927X13502793

Hess, N. H. & Hagen, E. H. (2006). Psychological adaptations for assessing gossip veracity. *Human Nature, 17,* 337–354. https://doi.org/10.1007/s12110-006-1013-z

Hobfoll, S. E. (1989). Conservation of resources. A new attempt at conceptualizing stress. *American Psychologist, 44*(3), 513–524. https://doi.org/10.1037/0003-066X.44.3.513

Hong, L. & Davison, B. D. (2010). Empirical study of topic modeling in Twitter. *SOMA '10: Proceedings of the First Workshop on Social Media Analytics* (pp. 80–88). Association for Computing Machinery. https://doi.org/10.1145/1964858.1964870

Husnu, S., Mertan, B. & Cicek, O. (2018). Reducing Turkish Cypriot children's prejudice toward Greek Cypriots: Vicarious and extended intergroup contact through storytelling. *Group Processes & Intergroup Relations, 21*(1), 178–192. https://doi.org/10.1177/1368430216656469

Jazaieri, H., Logli Allison, M., Campos, B., Young, R. C. & Keltner, D. (2019). Content, structure, and dynamics of personal reputation: the role of trust and status potential within social networks. *Group Processes & Intergroup Relations*, *22*(7), 964–983. https://doi.org/10.1177/1368430218806056

Jeuken, E., Beersma, B., ten Velden, F. S. & Dijkstra, M. T. (2015). Aggression as a motive for gossip during conflict: The role of power, social value orientation, and counterpart's behavior. *Negotiation and Conflict Management Research*, *8*(3), 137–152. https://doi.org/10.1111/ncmr.12053

Kazemzadeh, A., Lee, S. & Narayanan, S. (2013). Fuzzy logic models for the meaning of emotion words. *IEEE Computational Intelligence Magazine, 8*(2), 34–49. https://doi.org/10.1109/MCI.2013.2247824

Kisfalusi. D., Takács, K. & Pál. J. (2019). Gossip and Reputation in Adolescent Networks. In F. Giardini & R. P. M. Wittek (Eds.), *The Oxford Handbook of Gossip and Reputation* (pp. 359–379). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190494087.013.19

Koltai, J., Kmetty, Z. & Bozsonyi, K. (2021). From Durkheim to Machine Learning: Finding the Relevant Sociological Content in Depression and Suicide-Related Social Media Discourses. In T. Rudas & G. Péli (Eds.), *Pathways Between Social Science and Computational Social Science* (pp. 237–258). Springer.

Kurland, N. B. & Pelled, L. H. (2000). Passing the word: Toward a model of gossip and power in the workplace. *Academy of Management Review, 25*(2), 428–438. https://doi.org/10.2307/259023

Labonté, R. & Feather, J. (1996). *Handbook on Using Stories in Health Promotion Practice*. Health Canada.

Labov, W. & Fanshel, D. (1977). *Therapeutic Discourse*. Academic Press.

Levin, J. & Arluke, A. (1985). An exploratory analysis of sex differences in gossip. *Sex Roles, 12*(3–4), 281–286. https://doi.org/10.1007/BF00287594

Levy, K. E. & Franklin, M. (2014). Driving regulation: using topic models to examine political contention in the US trucking industry. *Social Science Computer Review*, *32*(2), 182–194. https://doi.org/10.1177/0894439313506847

Liu, J., Li, C., Carcioppolo, N. & North, M. (2016). Do Our Facebook Friends Make Us Feel Worse? A Study of Social Comparison and Emotion. *Human Communication Research, 42*(4), 619–640. https://doi.org/10.1111/hcre.12090

Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A. & Tingley, D. (2015). Computer-assisted text analysis for comparative politics. *Political Analysis, 23*(2), 254–277. https://doi.org/10.1093/pan/mpu019

Lyons, M. T. & Hughes, S. (2015). Malicious mouths? The Dark Triad and motivations for gossip. *Personality and Individual Differences, 78,* 1–4. https://doi.org/10.1016/j.paid.2015.01.009

Mandelbaum J. (2013). Storytelling in conversation. In J. Sidnell & T. Stivers (Eds.), *Handbook of Conversation Analysis* (pp. 492–508). Cambridge University Press.

Mangardich, H. & Fitneva, S. A. (2019). Gossip, Reputation, and Language. In F. Giardini & R. P. M. Wittek (Eds.), *The Oxford Handbook of Gossip and Reputation* (pp. 82–99). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190494087.013.5

Martinescu, E., Janssen, O. & Nijstad, B. A. (2019). Gossip and emotions. In F. Giardini & R. P. M. Wittek (Eds.), *The Oxford Handbook of Gossip and Reputation* (pp. 152–169). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190494087.013.9

McFarland, D. A., Ramage, D., Chuang, J., Heer, J., Manning, C. D. & Jurafsky, D. (2013). Differentiating language usage through topic models. *Poetics, 41*(6), 607–625. https://doi.org/10.1016/j.poetic.2013.06.004

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 262–272). Association for Computational Linguistics.

Mitra, T. & Gilbert, E. (2012). Have you heard? How gossip flows through workplace email. In *Proceedings of the Sixth International AAAI Conference on Web and Social Media* (pp. 242–249). AAAI Press. https://doi.org/10.1609/icwsm.v6i1.14260

Mohr, J.W. & Bogdanov, P. (2013). Introduction – Topic models: What they are and why they matter. *Poetics, 41*(6), https://doi.org/10.1016/j.poetic.2013.10.001

Németh, R., Katona, E. R. & Kmetty, Z. (2020). Az automatizált szövegelemzés perspektívája a társadalomtudományokban [Perspectives of automated text analysis in the social sciences]. *Szociológiai Szemle*, *30*(1), 44–62. https://doi.org/10.51624/SzocSzemle.2020.1.3

Nowak, M. A. & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature, 437*(7063), 1291–1298. https://doi.org/10.1038/nature04131

Paquette, J. A. & Underwood, M. K. (1999). Gender Differences in Young Adolescents' Experiences of Peer Victimization: Social and Physical Aggression. *Merrill-Palmer Quarterly, 45*(2), 242–266.

Pauw, L. S., Sauter, D. A., van Kleef, G. A. & Fischer, A. H. (2018). I hear you (not): Sharers' expressions and listeners' inferences of the need for support in response to negative emotions. *Cognition and Emotion, 33*(6), 1129–1143. https://doi.org/10.1080/02699931.2018.1536036

Pheko, M. M. (2018). Rumors and gossip as tools of social undermining and social dominance in workplace bullying and mobbing practices: A closer look at perceived perpetrator motives. *Journal of Human Behavior in the Social Environment, 28*(4), 449–465. https://doi.org/10.1080/10911359.2017.1421111

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H. & Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science, 54*(1), 209–228. https://doi.org/10.1111/j.1540-5907.2009.00427.x

Řehůřek, R. & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45–50). ELRA. https://doi.org/10.13140/2.1.2393.1847

Righi, S. & Takács, K. (2022). Gossip: Perspective Taking to Establish Cooperation. *Dynamic Games and Applications*. https://doi.org/10.1007/s13235-022-00440-4

Robbins, M. L. & Karan, A. (2020). Who Gossips and How in Everyday Life? *Social Psychology and Personality Science, 11*(2), 185–195. https://doi.org/10.1177/1948550619837000

Rosnow, R. L. (2001). Rumor and gossip in interpersonal interaction and beyond: A social exchange perspective. In R. M. Kowalski (Ed.), *Behaving badly: Aversive behaviors in interpersonal relationships* (pp. 203–232). American Psychological Association. https://doi.org/10.1037/10365-008

Schacter, D. L., Benoit, R. G. & Szpunar, K. K. (2017). Episodic Future Thinking: Mechanisms and Functions. *Current Opinion in Behavioral Sciences, 17,* 41–50. https://doi.org/10.1016/j.cobeha.2017.06.002

Samu, F. & Takács, K. (2021). Evaluating mechanisms that could support credible reputations and cooperation: cross-checking and social bonding. *Philosophical Transactions of the Royal Society B, 376*(1838), 20200302. https://doi.org/10.1098/rstb.2020.0302

Samu, F., Számadó, Sz. & Takács, K. (2020). Scarce and Directly Beneficial Reputations Support Cooperation. *Scientific Reports, 10,* 11486. https://doi.org/10.1038/s41598-020-68123-x

Shaikh, M. A. M., Prendinger, H. & Ishizuka, M. (2008). Sentiment assessment of text by analyzing linguistic features and contextual valence assignment. *Applied Artificial Intelligence, 22*(6), 558–601. https://doi.org/10.1080/08839510802226801

Shimanoff, S. B. (1985). Expressing Emotions in Words: Verbal Patterns of Interaction. *Journal of Communication, 35*(3), 16–31. https://doi.org/10.1111/j.1460-2466.1985.tb02445.x

Slade, D. (1997). Stories and gossip in English: The macro-structure of casual talk. *Prospect, 12*(2), 43–71.

Sommerfeld, R. D., Krambeck, H. J. & Milinski, M. (2008). Multiple gossip statements and their effect on reputation and trustworthiness. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1650), 2529-36. https://doi.org/10.1098/rspb.2008.0762

Suls, J. M. (1977). Gossip as social comparison. *Journal of Communication, 27*(1), 164–168. https://doi.org/10.1111/j.1460-2466.1977.tb01812.x

Szabó, M. K. (2014). Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai [Experiences of creation of a Hungarian sentiment lexicon]. "Nyelv, kultúra, társadalom" conference, Budapest, Hungary.

Szabó, M. K. & Morvay, G. (2015). Emócióelemzés magyar nyelvű szövegeken [Emotion analysis in Hungarian language texts]. In T. Gecső & Cs. Sárdi (Eds.). *Nyelv, kultúra, társadalom* (pp. 286–292). Tinta.

Szabó, M. K. & Vincze V. (2015). Egy magyar nyelvű szentimentkorpusz létrehozásának tapasztalatai [Experiences of creation of a Hungarian sentiment corpus]. In Tanács A., Varga V. & Vincze V. (Eds.), *XI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2015). Szegedi Tudományegyetem, Szeged* (pp. 219–226). https://acta.bibl.u-szeged.hu/58935/1/mszny konf_011_219-226.pdf

Szabó, M. K., Vincze, V. & Morvay, G. (2016). Magyar nyelvű szövegek emócióelemzésének elméleti és nyelvtechnológiai problémái [Theoretical and technical problems of emotion analysis of Hungarian texts]. In A. Á. Reményi, Cs. Sárdi & Zs. Tóth (Eds.), *Távlatok a mai magyar alkalmazott nyelvészetben* (pp. 282–292). Tinta.

Szabó, M. K., Vincze, V., Ring, O., Üveges, I., Vit, E., Samu, F., Gulyás, A., Galántai, J., Szvetelszky, Zs., Bodor-Eranus, E. H. & Takács, K. (2021). StaffTalk: magyar nyelvű spontán beszélgetések korpusza [StaffTalk: Corpus of spontaneous conversations in Hungarian]. In *XVII. Magyar Számítógépes Nyelvészeti Konferencia, Szegedi Tudományegyetem* (pp. 211–224).

Számadó, Sz., Balliet, D., Giardini, F., Power, E. A. & Takács, K. (2021). The language of cooperation: reputation and honest signalling. *Philosophical Transactions of the Royal Society B*, *376*(1838), 20200286. https://doi.org/10.1098/rstb.2020.0286

Tholander, M. (2003). Pupils' gossip as remedial action. *Discourse Studies, 5*(1), 101–128. https://doi.org/10.1177/1461456030050010501

Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology, 53*, 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Vancsó, A. & Kmetty, Z. (2021). Dominant Christian narratives of solidarity during the COVID pandemic in Hungary. *Intersections. East European Journal of Society and Politics*, *7*(3), 101–119. https://doi.org/10.17356/ieejsp.v7i3.804

Wert, S. R. & Salovey, P. (2004). A social comparison account of gossip. *Review of General Psychology*, *8*(2), 122–137. https://doi.org/10.1037/1089-2680.8.2.122

Wu, J., Balliet, D. & van Lange, P. A. (2015). When does gossip promote generosity? Indirect reciprocity under the shadow of the future. *Social Psychological and Personality Science*, *6*(8), 923–930. https://doi.org/10.1177/1948550615595272

Wu, J., Balliet, D. & van Lange, P. A. (2016a). Reputation, gossip, and human cooperation. *Social and Personality Psychology Compass*, *10*(6), 350–364. https://doi.org/10.1111/spc3.12255

Wu, J., Balliet, D. & van Lange, P. A. (2016b). Gossip versus punishment: The efficiency of reputation to promote and maintain cooperation. *Scientific Reports*, *6*, 23919. https://doi.org/10.1038/srep23919

Wu, J., Balliet, D. & van Lange, P. A. (2016c). Reputation management: Why and how gossip enhances generosity. *Evolution and Human Behavior, 37*(3), 193–201. https://doi.org/10.1016/j.evolhumbehav.2015.11.001

Zsibrita, J., Vincze, V. & Farkas, R. (2013). magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian. In *Proceedings of Recent Advances in Natural Language Processing* (pp. 763–771).

Appendix A

Ensuring quality

The quality of the transcription and annotation was tested in multiple dimensions and ensured in several ways. This included continuous supervision, qualitative control, automatized checks, and thorough checks of randomly selected samples. First, the annotators have been selected through a long trial process. Second, annotation marks and rules were developed in a trial phase using comments from and in interaction with annotators. Third, the quality of work of the annotators was measured by giving them the same texts and examining the transcripts' accuracy, the annotation tags, name tags, and timestamp usage, which are divided into sub-dimensions for more accurate feedback. We compared annotators by comparing their work and by using a reference annotator. Annotators with poor relative performance were suspended from further work and their texts were re-annotated when necessary. We provided individual feedback to annotators related to every quality assurance dimension.

These measures were continuously applied when the corpus was built. To compare our annotators' working quality, we used inter-coder agreement measures. Text similarity was measured by cosine similarity and Levenshtein distance. Due to the nature of spontaneous speech and the complexity of the annotation material we used indicators of text similarity and annotation agreement to evaluate annotators' performance. The inter-annotator agreement measures showed 74 per cent for participant (speaker) tags, 72 per cent for whether a text contained gossip tags, and 50 per cent for unique gossip tags after the deduplication of these tags for each row in the database (based on normed Levenshtein distances). The relatively low value in the last case indicates subjectivity bias in the evaluation of gossip.

Appendix B

Sentiment and emotion dictionaries

In order to measure the emotionality and sentiment characteristics of topics, besides the tagging of expressions and emotions we used a sentiment and an emotion dictionary developed and evaluated by the company Precognox (Szabó & Morvay, 2015; Szabó & Vincze, 2015). The sentiment dictionary consists of both positive and negative sentiments. Its reliability was investigated by measuring annotation agreement with two annotators at a 65.02 per cent

agreement rate (Szabó, 2014; Szabó, Vincze & Morvay, 2016). The emotion dictionary consists of six subcategories that rely on the Ekman-Friesen categories (Ekman & Friesen, 1969). It was translated based on the Affective Text Dictionary (Strapparava & Michalcea, 2007) and supplemented with additional synonyms. Later, for quality insurance purposes, the word matching method was used.

We used the dictionaries to identify basic correlations with the tags in our corpus and the topics obtained as a result of LDA topic modeling.

References

Ekman, P. & Friesen, W. V. (1969). The repertoire of nonverbal behavior: categories, origins, usage, and coding. *Semiotica, 1*(1), 49–98. https://doi.org/10.1515/SEMI.1969.1.1.49

Strapparava, C. & Mihalcea, R. (2007). SemEval-2007 Task 14: Affective Text. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)* (pp. 70–74). Association for Computational Linguistics. https://aclanthology.org/S07-1013

Szabó, M. K. (2014). Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai [Experiences of creation of a Hungarian sentiment lexicon]. "Nyelv, kultúra, társadalom" conference, Budapest, Hungary.

Szabó, M. K. & Morvay, G. (2015). Emócióelemzés magyar nyelvű szövegeken [Emotion analysis in Hungarian language texts]. In T. Gecső & Cs. Sárdi (Eds.). *Nyelv, kultúra, társadalom* (pp. 286–292). Tinta.

Szabó, M. K. & Vincze V. (2015). Egy magyar nyelvű szentimentkorpusz létrehozásának tapasztalatai [Experiences of creation of a Hungarian sentiment corpus]. In Tanács A., Varga V. & Vincze V. (Eds.), *XI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2015). Szegedi Tudományegyetem, Szeged* (pp. 219–226). http://acta.bibl.u-szeged.hu/58935/1/mszny konf_011_219-226.pdf

Szabó, M. K., Vincze, V. & Morvay, G. (2016). Magyar nyelvű szövegek emócióelemzésének elméleti és nyelvtechnológiai problémái [Theoretical and technical problems of emotion analysis of Hungarian texts]. In A. Á. Reményi, Cs. Sárdi & Zs. Tóth (Eds.), *Távlatok a mai magyar alkalmazott nyelvészetben* (pp. 282–292). Tinta.

Appendix C

Tables and Figures

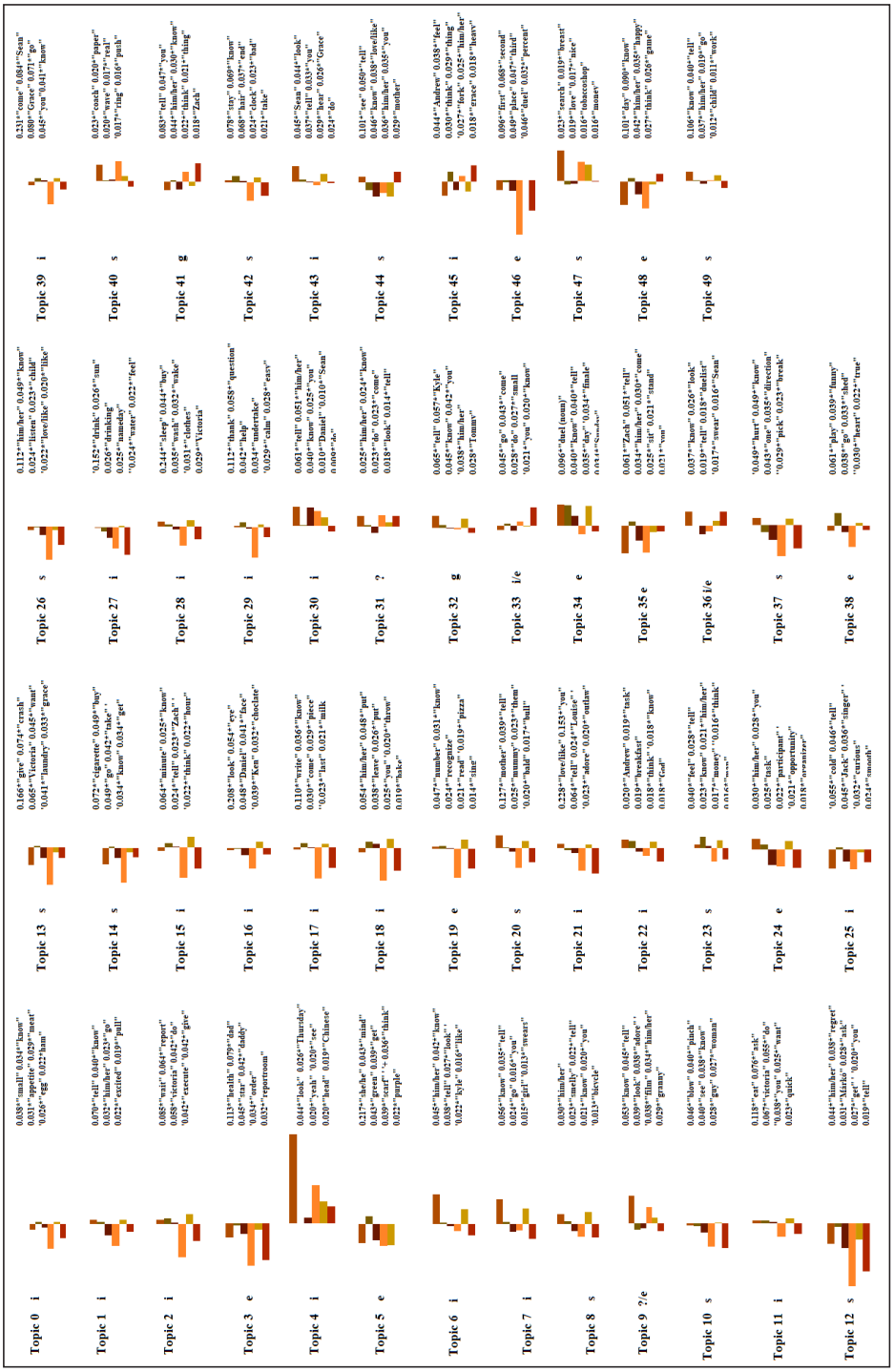

Figure 2 Topic summary by topic number.

Notes: Ex-post categorization: i = internal issues; e = entertainment show; s = storytelling; g = in-group gossip. Frequencies by colors:

Table 9 Topic summary by dictionaries and annotation tags, Spearman correlations (r_s) of topics and variables

	topic label	ratio of gossip tags		speakers present		participants present		number of words		ratio of joyful words		ratio of sadness		ratio of anger		ratio of fear		ratio of positive words		ratio of negative words		ratio of non-verbal tags	
	i	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p
topic_0	i	-0,064	***	0,011	0,078	0,081	***	-0,111	***	-0,016	0,007	0,016	0,008	0,008	0,21	0,024	***	-0,023	***	-0,02	0,001	-0,026	***
topic_1	i	-0,031	***	0,02	0,001	0,053	***	-0,092	***	-0,046	***	-0,003	0,579	0,01	0,1	0,021	0,001	-0,051	***	-0,045	***	0,02	0,001
topic_2	e	-0,072	***	0,042	***	0,082	***	-0,138	***	0,005	0,433	0,035	***	0,023	***	0,044	***	-0,008	0,178	-0,006	0,334	0,019	0,002
topic_3	i	-0,101	***	-0,016	0,008	0,042	***	-0,116	***	-0,029	***	0,005	0,392	-0,005	0,4	0,008	0,199	-0,035	***	-0,036	***	-0,039	***
topic_4	e	0,017	0,004	0,022	***	0,002	0,788	0,039	***	0,006	0,313	0,006	0,325	0	0,979	0,003	0,674	0,002	0,771	0,011	0,08	0,091	***
topic_5	i	0	0,989	-0,044	***	0	0,941	-0,045	***	-0,034	***	0,009	0,119	0,017	0,005	0,015	0,012	-0,04	***	-0,025	***	-0,039	***
topic_6	i	-0,05	***	0,064	***	0,075	***	-0,031	***	-0,01	0,093	0,023	***	0,004	0,465	0,02	0,001	-0,025	***	-0,036	***	0,126	***
topic_7	s	-0,062	***	0,065	***	0,074	***	-0,028	***	-0,034	***	0,011	0,061	0,008	0,195	0,021	***	-0,049	***	-0,039	***	0,105	***
topic_8	?/e	-0,056	***	0,051	***	0,076	***	-0,054	***	-0,031	***	0,003	0,608	0,013	0,03	0,017	0,004	-0,044	***	-0,03	***	0,042	***
topic_9	s	-0,033	***	0,025	***	-0,002	0,725	0,07	***	-0,02	0,001	-0,021	0,001	-0,026	***	-0,023	***	-0,026	***	-0,022	***	0,119	***
topic_10	i	-0,105	***	0,002	0,797	0,04	***	-0,1	***	-0,04	***	-0,012	0,058	-0,011	0,07	-0,002	0,784	-0,047	***	-0,04	***	-0,01	0,112
topic_11	s	-0,11	***	0,055	***	0,095	***	-0,142	***	0,013	0,034	0,04	***	0,028	***	0,044	***	0,007	0,262	-0,012	0,053	0,028	***
topic_12	s	-0,067	***	-0,023	***	0,027	***	-0,088	***	-0,035	***	0,003	0,567	-0,005	0,438	-0,001	0,868	-0,046	***	-0,035	***	-0,029	***
topic_13	s	-0,024	***	-0,011	0,079	0,049	***	-0,086	***	-0,025	***	0,002	0,689	0,003	0,649	0,011	0,073	-0,036	***	-0,027	***	-0,041	***
topic_14	i	-0,088	***	0,046	***	0,098	***	-0,127	***	0,001	0,888	0,03	***	0,018	0,003	0,03	***	-0,01	0,1	-0,016	0,007	-0,013	0,037
topic_15	i	-0,027	***	0,026	***	0,063	***	-0,086	***	-0,028	***	0,002	0,756	-0,004	0,544	0,013	0,033	-0,038	***	-0,037	***	-0,007	0,278
topic_16	i	-0,087	***	0,014	0,022	0,059	***	-0,13	***	-0,002	0,696	0,027	***	0,019	0,002	0,038	***	-0,006	0,355	-0,01	0,091	-0,008	0,179
topic_17	i	-0,095	***	0,038	***	0,075	***	-0,135	***	0,017	0,004	0,037	***	0,028	***	0,044	***	0,011	0,064	-0,001	0,933	-0,016	0,008
topic_18	e	-0,082	***	0,04	***	0,078	***	-0,123	***	0,003	0,588	0,035	***	0,014	0,018	0,043	***	-0,004	0,475	-0,013	0,027	0,009	0,119
topic_19	s	-0,06	***	0,031	***	0,059	***	-0,083	***	-0,013	0,037	0,017	0,006	0,003	0,661	0,016	0,008	-0,022	***	-0,026	***	0,055	***
topic_20	i	-0,107	***	0,017	0,005	0,057	***	-0,095	***	-0,02	0,001	0,004	0,478	-0,007	0,262	-0,001	0,822	-0,03	***	-0,033	***	0,018	0,003
topic_21	i	-0,055	***	0,028	***	0,07	***	-0,032	***	-0,013	0,033	0,029	***	0,03	***	0,032	***	-0,009	0,129	-0,025	***	0,038	***
topic_22	s	-0,047	***	0,035	***	0,097	***	-0,056	***	0,009	0,133	0,05	***	0,048	***	0,053	***	0	0,987	0,002	0,737	0,015	0,014
topic_23	e	-0,039	***	0,018	0,003	0,054	***	-0,036	***	-0,033	***	0,015	0,013	0,011	0,068	0,023	***	-0,041	***	-0,034	***	0,023	***
topic_24	i	-0,027	***	-0,006	0,317	0,066	***	-0,042	***	-0,026	***	0,005	0,446	0,004	0,487	0,014	0,021	-0,027	***	-0,027	***	-0,041	***
topic_25	s	-0,061	***	-0,011	0,068	0,031	***	-0,11	***	-0,028	***	0,013	0,039	-0,001	0,843	0,004	0,504	-0,038	***	-0,029	***	-0,011	0,066
topic_26	i	-0,08	***	0,005	0,375	0,049	***	-0,061	***	-0,03	***	0,001	0,813	-0,013	0,038	0,003	0,681	-0,048	***	-0,04	***	-0,002	0,785
topic_27	i	-0,058	***	0,026	***	0,054	***	-0,085	***	-0,013	0,031	0,011	0,06	0,01	0,118	0,027	***	-0,027	***	-0,018	0,004	0,021	***
topic_28	i	-0,045	***	0,013	0,038	0,04	***	-0,135	***	-0,006	0,348	0,025	***	0,021	0,001	0,039	***	-0,012	0,046	-0,018	0,003	0,005	0,404
topic_29	i	-0,006	0,323	0,009	0,136	0,006	0,327	0,016	0,008	0,02	0,001	0,002	0,745	0,001	0,848	-0,001	0,874	0,021	0,001	0,007	0,239	0,021	0,001
topic_30	?	0,117	***	0,045	***	0,031	***	0,125	***	-0,066	***	0	0,955	0,014	0,02	0,004	0,553	-0,08	***	-0,069	***	0,119	***
topic_31	g	-0,052	***	0,102	***	0,114	***	-0,011	0,08	-0,008	0,206	0,04	***	0,034	***	0,048	***	-0,026	***	-0,016	0,008	0,132	***
topic_32	i/e	0,202	***	-0,012	0,044	0,01	0,111	0,044	***	-0,058	***	-0,01	0,09	0,018	0,003	-0,008	0,172	-0,067	***	-0,058	***	-0,046	***
topic_33	e	-0,025	***	0,087	***	0,123	***	-0,039	***	0,043	***	0,088	***	0,09	***	0,107	***	0,029	***	0,024	***	0,092	***
topic_34	e	-0,013	0,027	-0,014	0,018	0,022	***	-0,06	***	-0,033	***	0,001	0,812	0,008	0,192	0,02	0,001	-0,042	***	-0,013	0,037	-0,061	***
topic_35	i/e	0,064	***	0,024	***	0,042	***	-0,024	***	-0,037	***	0	0,987	0	0,965	0,019	0,002	-0,039	***	-0,045	***	0,063	***
topic_36	s	-0,052	***	0,013	0,037	0,059	***	-0,069	***	-0,033	***	-0,007	0,277	-0,016	0,009	0,001	0,84	-0,042	***	-0,036	***	0,015	0,013
topic_37	e	-0,07	***	0,017	0,004	0,068	***	-0,09	***	-0,026	***	0,005	0,412	0,001	0,811	0,027	***	-0,033	***	-0,023	***	-0,014	0,026
topic_38	i	-0,01	0,097	0,005	0,45	0,059	***	-0,046	***	-0,014	0,021	0,017	0,005	0,027	***	0,028	***	-0,023	***	-0,02	0,001	-0,011	0,079
topic_39	s	-0,048	***	0,024	***	0,068	***	-0,145	***	0,011	0,067	0,032	***	0,026	***	0,043	***	0,005	0,39	-0,009	0,151	-0,019	0,002
topic_40	g	-0,019	0,001	0,017	0,005	0,006	0,365	0,067	***	0,005	0,431	0,011	0,075	0,003	0,576	0,012	0,041	0,003	0,62	0,006	0,353	0,055	***
topic_41	s	0,235	***	-0,051	***	-0,041	***	0,132	***	-0,091	***	-0,037	***	0,024	***	-0,04	***	-0,092	***	-0,072	***	-0,105	***
topic_42	i	-0,065	***	0,025	***	0,057	***	-0,093	***	0,009	0,13	0,035	***	0,032	***	0,043	***	0,001	0,849	0,002	0,693	0,012	0,042
topic_43	s	-0,019	0,002	0,096	***	0,112	***	-0,047	***	-0,012	0,046	0,033	***	0,026	***	0,041	***	-0,029	***	-0,024	***	0,188	***
topic_44	i	0,027	***	-0,037	***	-0,001	0,851	-0,028	***	-0,036	***	0,009	0,134	-0,021	***	-0,002	0,761	-0,045	***	-0,051	***	0,014	0,018
topic_45	e	0,042	***	-0,026	***	0,014	0,026	0,015	0,013	-0,022	***	0,012	0,044	0,025	***	0,027	***	-0,031	***	-0,028	***	-0,038	***
topic_46	s	-0,053	***	0	0,995	0,046	***	-0,095	***	-0,019	0,002	0,002	0,759	-0,004	0,524	0,023	***	-0,027	***	-0,026	***	-0,017	0,006
topic_47	e	-0,003	0,646	0,04	***	0,024	***	0,048	***	-0,008	0,217	-0,006	0,333	-0,01	0,097	-0,006	0,302	-0,011	0,059	-0,015	0,017	0,076	***
topic_48	s	0,01	0,103	-0,003	0,603	0,032	***	-0,034	***	-0,016	0,011	0,004	0,534	0,005	0,45	0,023	***	-0,031	***	-0,028	***	-0,029	***
topic_49		-0,09	***	0,072	***	0,063	***	0,001	0,824	-0,035	***	0,008	0,213	0,004	0,54	0,034	***	-0,05	***	-0,038	***	0,114	***

Table 9 *(Continuing)*

	coughing		sighing		laughter		crying		mocking laughter		confused laughter		screaming		yawning		hiss		exultation		throat clearing	
	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p
topic_0	0,042	***	0,017	0,006	-0,042	***	0,033	***	0,023	***	0,014	0,019	0,034	***	0,026	***	0,034	***	0,037	***	0,016	0,008
topic_1	0,037	***	0,035	***	0,005	0,436	0,032	***	0,009	0,123	0,023	***	0,028	***	0,014	0,026	0,012	0,045	0,012	0,048	0,018	0,003
topic_2	0,079	***	0,059	***	0,005	0,41	0,033	***	0,049	***	0,06	***	0,071	***	0,064	***	0,059	***	0,056	***	0,03	***
topic_3	0,021	0,001	-0,018	0,004	-0,045	***	0,022	***	-0,018	0,003	-0,012	0,046	0,01	0,093	0,006	0,308	0,021	0,001	0,015	0,014	-0,024	***
topic_4	0,037	***	0,028	***	0,09	***	0,004	0,502	0,043	***	0,027	***	-0,005	0,389	0,014	0,025	0	0,99	0,015	0,014	0,033	***
topic_5	0,008	0,215	0,009	0,125	-0,054	***	0,03	***	-0,01	0,111	0	0,949	-0,006	0,31	0,013	0,027	0,027	***	0,011	0,066	0,047	***
topic_6	0,082	***	0,054	***	0,134	***	0,02	0,001	0,064	***	0,062	***	0,045	***	0,057	***	0,042	***	0,042	***	0,037	***
topic_7	0,093	***	0,055	***	0,104	***	-0,003	0,637	0,057	***	0,075	***	0,049	***	0,043	***	0,018	0,003	0,031	***	0,051	***
topic_8	0,06	***	0,037	***	0,039	***	0,011	0,061	0,028	***	0,015	0,017	0,02	0,001	0,025	***	0,03	***	0,021	0,001	0,051	***
topic_9	0,021	0,001	0,004	0,524	0,123	***	-0,031	***	0,042	***	0,008	0,167	-0,005	0,412	-0,004	0,558	-0,027	***	-0,027	***	0,01	0,088
topic_10	0,005	0,417	0,009	0,121	-0,011	0,084	-0,001	0,855	-0,005	0,409	-0,006	0,299	0,004	0,532	0,003	0,592	0,006	0,36	-0,009	0,146	-0,001	0,872
topic_11	0,072	***	0,046	***	0,013	0,034	0,073	***	0,028	***	0,03	***	0,054	***	0,058	***	0,058	***	0,068	***	0,022	***
topic_12	0,001	0,847	-0,003	0,59	-0,032	***	-0,004	0,505	-0,007	0,237	-0,015	0,014	-0,002	0,701	0,009	0,12	0,009	0,154	-0,005	0,392	0,018	0,003
topic_13	0,04	***	0,012	0,052	-0,044	***	0,015	0,017	0,001	0,871	0	0,987	0,002	0,723	0,02	0,001	0,02	0,001	0,017	0,006	0,014	0,026
topic_14	0,057	***	0,042	***	-0,027	***	0,034	***	0,038	***	0,032	***	0,05	***	0,054	***	0,049	***	0,056	***	0,014	0,021
topic_15	0,032	***	0,023	***	-0,019	0,002	0,014	0,026	0,018	0,004	0,013	0,037	0,014	0,024	0,01	0,09	0,014	0,02	0,024	***	0,018	0,003
topic_16	0,05	***	0,016	0,007	-0,002	0,735	0,033	***	0,017	0,006	0,032	***	0,071	***	0,053	***	0,053	***	0,051	***	0,007	0,237
topic_17	0,065	***	0,045	***	-0,026	***	0,038	***	0,037	***	0,021	***	0,049	***	0,06	***	0,07	***	0,06	***	0,012	0,049
topic_18	0,042	***	0,032	***	0,003	0,573	0,021	0,001	0,025	***	0,023	***	0,049	***	0,055	***	0,051	***	0,039	***	0,009	0,145
topic_19	0,049	***	0,032	***	0,04	***	0,047	***	0,048	***	0,027	***	0,058	***	0,023	***	0,036	***	0,036	***	0,028	***
topic_20	0,033	***	0,001	0,912	0,022	***	0,013	0,034	0,021	0,001	-0,001	0,909	0,014	0,018	0,027	***	0,014	0,026	0,014	0,023	0,006	0,334
topic_21	0,036	***	0,043	***	0,024	***	0,057	***	0,04	***	0,048	***	0,043	***	0,052	***	0,048	***	0,045	***	0,053	***
topic_22	0,069	***	0,045	***	0,009	0,156	0,055	***	0,053	***	0,036	***	0,066	***	0,07	***	0,08	***	0,072	***	0,037	***
topic_23	0,024	***	0,037	***	0,01	0,099	0,03	***	0,038	***	0,042	***	0,031	***	0,025	***	0,027	***	0,012	0,054	0,052	***
topic_24	0,023	***	0,039	***	-0,058	***	0,034	***	-0,001	0,927	0,007	0,219	0,023	***	0,013	0,034	0,016	0,008	0,033	***	0,025	***
topic_25	0,038	***	0,017	0,005	-0,024	***	0,017	0,005	-0,004	0,504	-0,007	0,239	0,011	0,081	0,034	***	0,013	0,028	0,019	0,002	0,021	0,001
topic_26	0,021	0,001	-0,002	0,741	-0,016	0,009	0,008	0,202	0,009	0,121	-0,008	0,177	0,001	0,84	0,011	0,073	0,006	0,33	-0,002	0,747	0,011	0,076
topic_27	0,041	***	0,023	***	0,014	0,024	0,037	***	0,008	0,177	0,012	0,054	0,032	***	0,026	***	0,033	***	0,043	***	0,031	***
topic_28	0,033	***	0,05	***	-0,009	0,147	0,051	***	0,03	***	0,032	***	0,039	***	0,051	***	0,045	***	0,03	***	0,019	0,002
topic_29	0,019	0,002	0,024	***	0,007	0,254	0,007	0,279	-0,01	0,092	0,001	0,926	0,005	0,414	0,001	0,824	0,001	0,9	0	0,945	0,003	0,6
topic_30	0,044	***	0,061	***	0,107	***	0,001	0,877	0,043	***	0,057	***	-0,001	0,844	0,014	0,025	-0,011	0,084	-0,014	0,018	0,058	***
topic_31	0,075	***	0,08	***	0,134	***	0,034	***	0,069	***	0,06	***	0,104	***	0,076	***	0,055	***	0,063	***	0,074	***
topic_32	-0,018	0,003	0,022	***	-0,073	***	0,004	0,558	-0,004	0,505	0,007	0,27	-0,012	0,051	-0,011	0,07	-0,01	0,118	0,002	0,745	0,037	***
topic_33	0,106	***	0,107	***	0,083	***	0,115	***	0,107	***	0,113	***	0,132	***	0,107	***	0,128	***	0,144	***	0,078	***
topic_34	0,013	0,033	0,014	0,021	-0,073	***	0,026	***	-0,019	0,002	-0,011	0,081	-0,001	0,829	0,015	0,015	0,019	0,002	0,016	0,008	0,028	***
topic_35	0,053	***	0,068	***	0,041	***	0,064	***	0,039	***	0,052	***	0,043	***	0,019	0,002	0,006	0,316	0,042	***	0,05	***
topic_36	0,033	***	0,031	***	0,002	0,803	0,005	0,437	-0,002	0,776	-0,006	0,356	0,023	***	-0,001	0,82	-0,003	0,568	0,01	0,115	0,015	0,015
topic_37	0,043	***	0,022	***	-0,022	***	0,046	***	0,02	0,001	0,023	***	0,033	***	0,026	***	0,025	***	0,042	***	0,007	0,223
topic_38	0,011	0,069	0,023	***	-0,025	***	0,05	***	0,023	***	0,032	***	0,04	***	0,018	0,003	0,04	***	0,049	***	0,042	***
topic_39	0,061	***	0,053	***	-0,031	***	0,071	***	0,026	***	0,026	***	0,067	***	0,057	***	0,058	***	0,068	***	-0,003	0,653
topic_40	0,021	0,001	0,015	0,015	0,054	***	0,016	0,007	0,019	0,002	0,014	0,026	0,002	0,743	0,001	0,881	0,009	0,138	-0,004	0,505	0,007	0,238
topic_41	-0,055	***	-0,01	0,106	-0,123	***	-0,048	***	-0,034	***	-0,007	0,241	-0,062	***	-0,056	***	-0,038	***	-0,068	***	0,016	0,008
topic_42	0,05	***	0,026	***	0,007	0,253	0,065	***	0,039	***	0,049	***	0,066	***	0,049	***	0,066	***	0,056	***	0,025	***
topic_43	0,111	***	0,077	***	0,189	***	0,047	***	0,102	***	0,108	***	0,12	***	0,06	***	0,049	***	0,067	***	0,047	***
topic_44	0,021	***	0,026	***	0,001	0,915	0,078	***	-0,007	0,267	0,021	0,001	0,009	0,134	-0,016	0,01	0,003	0,59	0,01	0,092	0,015	0,011
topic_45	0,018	0,002	0,048	***	-0,061	***	0,037	***	-0,005	0,388	0,022	***	0,022	***	0,02	0,001	0,039	***	0,03	***	0,036	***
topic_46	0,039	***	0,028	***	-0,026	***	0,04	***	-0,006	0,339	0,016	0,01	0,045	***	0,018	0,003	0,023	***	0,046	***	0,024	***
topic_47	-0,004	0,564	0,019	0,001	0,069	***	-0,007	0,252	0,014	0,021	0,024	***	0,014	0,019	-0,016	0,008	-0,02	0,001	0,001	0,857	0,012	0,041
topic_48	0,016	0,009	0,03	***	-0,042	***	0,029	***	0,021	0,001	0,045	***	0,024	***	0,021	0,001	0,024	***	0,033	***	0,034	***
topic_49	0,062	***	0,045	***	0,12	***	0,016	0,008	0,074	***	0,055	***	0,037	***	0,034	***	0,017	0,004	0,006	0,327	0,068	***

Topics for illustration that emerged as a result of an LDA topic model based on a document-term matrix with 12,961 documents (segments) and 8,530 terms, df=12959. *** significance level is $p < 0.001$.

Funding

The project was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program grant agreement No 648693, PI: Károly Takács.

Acknowledgments

We thank for permitting usage of the MTA Cloud (https://cloud.mta.hu/) of the the Hungarian Academy of Sciences, where the corpus has been stored and analysis was conducted. We are grateful to Precognox for allowing us to use their dictionaries. We thank Martina Szabó, Flóra Samu, Eliza Bodor-Eranus, and participants of the Text2Story workshop (Grenoble 2018) and the Gossip, Reputation, and Honesty workshop (Budapest 2018) for their comments and help.