

Textual Analysis for Radicalisation Narratives aligned with Social Sciences Perspectives

Ronald Denaux
Expert System, Madrid, Spain,
rdenaux@expertsystem.com

Jose Manuel Gómez-Pérez
Expert System, Madrid, Spain,
jmgomez@expertsystem.com

Abstract

One of the unintended consequences of the Web is that it can function as a radicalising medium. Hence, developing information systems that are capable of detecting radicalising content is one of the key challenges faced by society to prevent and minimise radicalisation. Fortunately, much work has already been done by social scientists to understand key factors in the radicalisation process and common narratives. This paper presents work to reuse this understanding from social science in a way that is useful for designing and developing information systems. We present work summarising various perspectives on the concept of narratives and how they apply to radicalisation domains; in particular, we focus on islamic radicalisation as a key example of radicalisation. We introduce three taxonomies to help capture different aspects of radicalisation narratives and present a system for identifying mentions of one of such aspects in texts: strategic radicalisation narratives for islamic radicalisation.

1 Introduction

In recent years, there is a realisation that the Web, and more specifically Social Media combined with viral recommendations, can function as a radicalising medium; especially for vulnerable sections of the population (young, elderly, minorities and isolated people). Radical groups are exploiting this medium to communicate their message and recruit new members. In particular, various islamic radical groups like ISIS and Al-Qaeda have well-documented strategies for targeting and recruiting members via social media [GRBM16]. This calls for automated systems to aid human operators to detect and analyse radicalising content.

Traditional NLP approaches have been applied to tackle the problem of automatically detecting radicalising content on-line. However, simple keyword-based and machine learning approaches tend to have poor performance due to its reliance on keywords and poor quality training datasets¹. More advanced machine learning approaches use additional information such as semantics [SDK⁺17] and social-network signals [FAA18]. In particular this last work started looking at social science models of the radicalisation process to inform the design of their system, enabling them to associate social-media signals with radicalisation factors. In their literature review, it becomes apparent that practically none of the existing approaches for analysis, detection and prediction of radicalisation reuse existing Social Science Models (an exception is [LCGPC17] which focuses on social media and proposes a keyword-based approach).

Copyright © 2019 for the individual papers by the paper's authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

In: A. Jorge, R. Campos, A. Jatowt, S. Bhatia (eds.): Proceedings of the Text2StoryIR'19 Workshop, Cologne, Germany, 14-April-2019, published at <http://ceur-ws.org>

¹<http://www.bbc.co.uk/newsbeat/article/34919781/anonymous-anti-islamic-state-list-features-obama-and-bbc-news>

Social scientists have been studying and modelling radicalisation processes for decades and have developed models to think about both the radicalisation process and communication strategies that radical groups use in order to convey their message and recruit new members [KT11, Du 17, GRBM16]. In this paper we argue that much of this work can and should be used to inform the design and development of radicalisation detection systems. In particular, most of the existing systems relying on machine-learning approaches behave as black-box systems and are therefore difficult to use and trust in practice. Being able to refer to existing radicalisation models can help to explain classification results of such systems.

In this paper, we look into how we can adapt notions of radicalisation narratives from social sciences into NLP systems. The main challenges we tackle are: (i) ill defined concept of narrative; (ii) mismatch between human and textual analysis capabilities; (iii) lack of formal, machine-actionable, representations for narrative analysis; (iv) lack of human annotated content to train automated detection systems and (v) unbalanced appearance of narratives in real-world texts.

Our main contributions are: (i) identification of three types of narratives from social science in Sect. 2, (ii) a set of taxonomies aligned to social science models (described in Sect. 3), (iii) an implementation of one of these taxonomies as a text annotation system and an analysis of various magazines published by islamic radical groups in Sect. 4.

2 Social Science Perspectives on Narratives

Social scientists and domain experts tend to use the term *radical narrative* loosely to *describe the main messages that are disseminated by a radical organisation*. In order to implement automated systems, we need to have a better understanding of what a narrative is. Performing an extensive literature review on the conceptualisation of this term is not in the scope of this paper, however we review a few representative works that provide a wider understanding of the term that can help us to focus the scope and hence, help us to generate useful semantic resources for radicalisation narratives.

From a generic, human sciences perspective, Riessman [Rie05] states that any text can be narrative if in it: “events are selected, organised, connected, and evaluated as meaningful for a particular audience”. Riessman introduces 4 main types of narrative analysis: (i) **thematic analysis** focuses on what is said in the text; (ii) **structural analysis** focuses on how the message is structured; (iii) **interactional analysis** looks at how the teller interacts with the listener and (iv) **performative analysis** looks at other factors besides the spoken or written word such as actions performed by the teller alongside the messages. In the context of this paper, we will focus mainly on supporting thematic narrative analysis with some basic structural analysis. Technological approaches for aiding in interactional and performative analyses are not in the scope of this paper.

Lewandowsky et al. [LSF⁺13] provide a *psychological point of view* for narratives in the context of conflicts and disinformation[Com18]. They consider narratives as mental frames, i.e. “necessary cognitive tools, designed to pare down information in order to manage complexity”, hence they are not the same as propaganda or spin; rather they facilitate communication. However, due to their inherent (over)simplification of reality, narratives can be misused to spread misinformation more effectively by making the misinformation coherent with the prevalent narrative. Thus, the *prevalence* of a narrative, i.e. whether alternative narratives are ignored by the media or the recipients, is important when considering its potential misuse.

Focusing more on radicalisation, and more specifically Islamic extremism, Halverson et al. [HGC11] defines a narrative not as a single story, but as a *system of stories*; i.e. a collection that is coherent and reinforces the same themes. When this definition of narrative is further linked to a (cultural) group identity, it becomes a master narrative. For example, for Muslims, the sacred texts provide a collection of stories (i.e. a narrative) that tells them who they are and how they should behave. The book [HGC11] further describes 11 “master narratives” employed by Islamist extremists to “connect or resonate within a set of cultural and historical circumstances”. It is worth noting that the individual narratives on their own are not extremist, they simply collect various versions of stories about specific significant events, periods and/or myths in the history of islam; however, Islamist extremists refer to these to increase their appeal to muslims.

Betz [Bet08] focuses on narratives in the context of conflicts, also called **strategic narratives** and uses a definition by Sir Lawrence Freedman as “compelling storylines which can explain events convincingly and from which inferences can be drawn”. These narratives are deliberately constructed out of current ideas, express a “sense of identity and belonging” and provide a “sense of purpose”. Importantly, these narratives often “appeal to emotion” and make use of “suspect metaphors and dubious historical analogies”. [Bet08] also introduces the concept of *vertical narrative coherence*, which states that there are multiple levels of narratives and that

successful strategic narratives are those that manage to coherently link to both master and individual narratives. These are the most useful to use as a basis for developing automated systems:

- **Cultural narratives** provide a cultural identity and are grounded in tradition. The master narratives described in [HGC11] are an example of this type of narrative. This also maps to the mass-level radicalisation indicators [Du 17, FAA18].
- **Strategic narratives** provide a group identity and are grounded in ideology and map to the meso level radicalisation indicators.
- **Local and individual narratives** are stories of specific places and individuals; these map to micro-level radicalisation indicators.

3 Radicalisation Narrative Taxonomies

Based on the three types of narratives described in the previous section, we set out to define taxonomies that can be used as the basis of text annotation tasks, in particular to aid in the detection of radical texts. One challenge in this regard is that in general, *it is not possible to define a single taxonomy that is generally applicable to all types of radicalisation* (e.g. radical islam, left- and right-wing extremism, white nationalism). Generalisability in this sense can apply to the taxonomy itself but also to the text analyser built to annotate texts according to the taxonomy. Of course, if the taxonomy itself is not generalisable, then the annotator targeting it is also not generalisable. The main reason impeding generalisability is that, as we saw in the previous section, the various narratives depend on specific knowledge about the cultural and radical group as well as on the individual.

- The **cultural narratives** are the *least generalisable*: it is necessary to derive custom taxonomies for each cultural group; e.g. cultural narratives about the *Battle of Badr*² or the *72 virgins*³ are specific to a muslim cultural identity, but not to e.g. a christian or jewish cultural identity (although both narratives can have similar parallels in those cultures). Fortunately, the number of cultural identities is fairly limited and lists of narratives are relatively easy to acquire.
- **Strategic narratives** are *generalisable up to a point*: the main strategies used by radical groups are generalisable, but the specifics can only be captured by taking into account the characteristics of the radical group. With other words, the main categories in the taxonomy are generally applicable, but categories deep in the taxonomy may be necessary to model strategic narratives used by –and applicable only– to specific radical groups. This lack of generalisability also applies to text annotators targeting even intermediate levels in the taxonomy. For example, a general strategy is to *discredit other groups*, but if you want to capture this in more detail to analyse whether the group being discredited is perceived to be *competition* or *enemy*, then a single article published by the radical group on its own is often insufficient (it may be possible to infer such relations from a larger corpus, but this is out of the scope of this paper). In such cases you need to add domain knowledge into the annotator, e.g. ISIS considers Al-Qaeda and the Muslim Brotherhood to be a competitor, while they consider e.g. christians and *the West* as enemies. Similarly, radical groups, especially online, often develop their own jargon to refer to aspects of their group identity. For example, radical islamist use the term *kufr* as a despective term to refer to non-believers. Such non-generalisable knowledge is even more prevalent when trying to capture aspects of a group’s identity. For example, *Incel* group identity relies on a large number of terms⁴ to refer to subgroups; hence knowing the correct term to refer to the most radical members in that group again requires custom knowledge tailored to the incel group identity. Such domain knowledge can be injected into annotation systems using different approaches like rule-based (explored in this paper) or machine-learning based (not in the scope of this paper, but relying on the annotation of large corpora).
- **Individual narratives** are the *most generalisable* as individuals tend to follow similar radicalisation processes and descriptions of these can be captured by looking for mentions of radicalisation indicators [Du 17, FAA18] which are not specific to a radicalisation type: grievances, meeting radicalised people, personal circumstances (death of a relative, unemployment, history of crime). These narratives can

²http://trivalent.expertsystemlab.com/thes/conceptschemes/ISLAMIC_MASTER_NARRATIVE/c/3

³http://trivalent.expertsystemlab.com/thes/conceptschemes/ISLAMIC_MASTER_NARRATIVE/c/12

⁴<https://incels.wiki/w/Incelese>

be found in stories and interviews about how someone was radicalised. Such articles typically use standard language (not jargon specific to the radical group), therefore it should be possible to build annotation systems which are capable of finding such mentions in a fairly generic way. However, one problem with building this type annotation systems is that there are not many available documents containing this type of narratives; also, detecting such stories is not directly useful for detecting radical texts or preventing radicalisation.

In this paper we aim to improve the automatic support provided to human analysts by defining taxonomies that can be implemented as automated text annotation systems that facilitate radicalisation narrative analysis. To this end, we present a taxonomy of Islamist extremist narratives which takes into account the three types of narratives discussed above. The full taxonomy has 89 nodes organized in 3 main sub-taxonomies⁵:

- **Islamic Master Narrative:** contains 37 subcategories, divided in 3 sublevels. The 11 subcategories in the first sublevel are those proposed in [HGC11] and further sublevels provide more specific subarguments within the narratives.
- **Strategic Narrative:** contains 20 subcategories, divided in 3 sublevels. The first 2 sublevels are *generic narratives* that can be exploited to promote radical ideologies by any radical group, and the 3rd sublevel is specific to strategic narratives by muslim extremist groups, in particular ISIS. The taxonomy has been derived from two lists of ISIS narratives, one presented in [GRBM16] and another provided by domain experts at the International Institute for Counter-Terrorism⁶. The first (generic) sublevel consists of categories:
 - **(promote) group identity:** any narrative that promotes the radical group spreading this message, e.g. by promoting a winner’s image for the group, promoting the social aspects of the group, associating the group with a sense of adventure and in general promoting the ideology of the group (or encompassing groups).
 - **Discredit other groups:** any narrative that discredits groups other than the group spreading the message. Typically the enemy is attacked, although similar but competing groups are often also discredited.
 - **Sow discord between groups:** whereby divisions between groups are exploited and reinforced.
 - **Moral obligation:** narratives that appeal to social and cultural norms to justify a certain type of action.
- **Radicalisation indicator:** contains 28 subcategories, divided in 2 sublevels, referring to structural causes and trigger events at all 3 (macro, meso and micro) levels or the radicalisation process. All of these subcategories are generic and not specific to a specific type of radicalisation. The taxonomy itself does not define subcategories for structural cause, trigger event, macro-, meso- or micro-level; however all of the subcategories are linked to a radicalisation ontology, which can be used to automatically classify the subcategory into one of these axes. This taxonomy was derived from the narrative conceptualisations described in [Du 17, KT11].

4 Implementation and Detection Results

We have implemented a rule-based system that annotates **Strategic Narratives** in texts. These rules are written in a custom rule language on top of Expert System’s NLP pipeline. The core of the pipeline consists of standard tokenization and lemmatization steps and a unique disambiguation step, which links n-grams in the text to concepts in a lexico-semantic knowledge graph (called Sensigrafo, which is similar to WordNet). This pre-analysed content is then fed into our rule engine, which matches parts of the text to rules hand-crafted by linguists to identify segments of text that should be annotated with one of the nodes in the taxonomy. Rules may refer to lemmas, keywords, concepts and related concepts (by following specific links in the knowledge graph) and can be combined using a variety of positional, syntactic and semantic operators. The resulting categorizer uses 80 rules, some of which rely on recognising words and concepts that are specific to muslim extremist slang.

In order to evaluate the rule-based narrative detector, we analysed a few collections of magazines:

⁵Available from <http://trivalent.expertsystemlab.com/thes/>

⁶<http://www.ict.org.il/>

- as a baseline, we selected 9 articles from “Young Muslim Digest”⁷ (YMD), a long-running magazine from India which is **non-radical**. These articles were selected from their homepage as well as from the “popular” articles recommendations from their *about* page. Topics are centred around lifestyle choices influenced by islam (e.g. husband-wife relationships, youth, conversion to islam), but also geo-politics (e.g. mining opportunities in Afghanistan, war).
- three issues of *Al-Risalah*, published by the Nusra Front[Kot18]
- 14 English issues of *Dabiq*, published by ISIS between 2014 and 2016
- 9 issues of of Rumiayah, a magazine published by ISIS since mid 2016

Table 1: Percentage of pages showing a strategic radicalisation narrative in various publications.

strategic narrative	% pages			
	ymd	risalah	dabiq	rumiyah
Sow discord between groups	0	5.56	19.95	23.79
Moral obligation	1.89	0	1.17	7.66
Protect your group	0	6.94	1.63	2.42
Avenge your group	0	0	2.80	0.40
Attack as self-defense	0	1.39	1.52	0.40
Group identity	0	0	0.35	7.26
Winners image	0	0	0	0
Winner in difficult times	0	1.39	0.47	0.40
Group’s achievements	0	1.39	6.42	16.94
Promote group ideology	1.89	2.78	4.20	5.65
Legitimacy of ideology	37.74	43.06	39.32	63.31
Everlasting essence of caliphate	0	0	0.82	0.40
Caliphate as islamic utopia	0	0	0.12	0.00
Homophily	9.43	65.28	28.70	36.29
Idealize most radical members	0	9.72	11.67	29.03
Glorify subgroups	0	0	0.70	0.81
Adventure and resilience	0	0	0	0
Discredit other group	0	0	32.79	58.47
Discredit enemy	3.77	65.28	45.62	57.66
Discredit West	0	0	0.23	0.40
Discredit christians in muslim lands	0	0	0.58	0.40
Discredit competition	0	0	0	0
Discredit political islamists	0	6.94	26.25	27.02
Discredit other jihadist groups	0	0	0	0

Table 1 provides the results which, overall, show (i) a stark contrast between the radical publications and the non-radical baseline (YMD) and (ii) diverging prevalence of narratives.

The **Group Identity** narrative is the most complex branch in the taxonomy, but shows clear differences between the publications. The only topic that has a relatively high coverage percentage in the non-radical texts was *legitimacy of ideology*. Manual inspection shows that this is caused by matching of sentences of the form “X says in the *Qur’an*: ...”, where X is an imam, important person or deity. Indeed, YMD uses this type of phrases frequently to legitimize lifestyle choices they recommend to their readers.

Homophily also occurs in YMD (in about 10% of the pages), but is much more prevalent in the radical publications. Furthermore, while YMD uses only generic homophily phrases (e.g. “*brothers and sisters*”), the radical publications use terminology to glorify and idealize the more radical members (e.g. “*martyrs*”). Radicalising publications also push a *winner* narrative as part of their group identity, especially by highlighting perceived *achievements*.

⁷<http://www.youngmuslimdigest.com>

resulted in approximately 100 murtaddin being killed and wounded.

Idlib

On the 5th of Muharram, our brother Abu Qudamah ash-Shami _ carried out an istishhadi attack, on a large group of American-vetted Sahwah fighters seeking to enter through the Atimah border crossing in order to fight the Islamic State. He detonated his explosive vehicle in their midst, killing 35 murtaddin and injuring nearly 45 others. Among them were some heads of kufr, including the murtadd Hisham Khalifah, a leader in Ahrar ash-Sham, the

Node / Rules	Description	Score	%
01.02.04.01	Group's achievements	1020	87.78 %
■ attack was carried out		80	
■ detonated his explosive ...		130	
■ killing and injuring X m...		540	
■ leading to many of thei...		270	
01.02.02.02	Idealize most radical ...	50	4.30 %
01.02.02	Homophily	40	3.44 %
01.01	Discredit other group	40	3.44 %
01.01.02	Discredit enemy	6	0.52 %
01.04	Sow discord between ...	3	0.26 %
01.01.01.02	Discredit political isla...	3	0.26 %

Figure 1: Example result for narrative “Group’s achievements” on page 43 of Rumiya issue 3. Because the rules are hand-crafted and refer to concepts in the Knowledge Graph, it is possible to highlight the segment that matched a rule as well as words and concepts. On the left we see the rules associated to this narrative that matched a phrase in this page.

By far the biggest difference between radical and non-radical texts is that radical texts actively push narratives to **discredit other groups**, in particular groups seen as the enemy or as ineffective alternatives (e.g. political islamists). This is closely related to the narrative of sowing discord between groups, where we see a similar pattern.

Another interesting result is that while YMD also includes narratives of **moral obligations**, these are framed in a neutral manner. By contrast, radical publications frame them in terms relative to the group: i.e. as an imperative to *protect, avenge or pre-emptively attack*. Note that, because the annotator is rule-based, it is straightforward to provide explanations for the annotations as shown in Figure 1.

5 Conclusion and Future Work

In this paper, we presented an approach for analysing radicalisation narratives in a way that is aligned with existing social science approaches. This resulted in a set of three taxonomies that can be used to annotate radicalisation narratives at the cultural, strategic and personal level. We also presented a rule-based prototype to automatically detect strategic narratives for islamic publications and showed that the results align with our expectations, can easily be explained, can be used to distinguish between radical and non-radical texts, in particular to identify prevalent narratives [LSF⁺13].

A key challenge in automating this type of narrative analysis is that it is crucial to encode knowledge about the originator of the text (the radical groups), their slang and who is the (intended) recipient; all of which impacts generalisability, as discussed in Section 3. Our results form a first step toward constructing a line of argumentation based on how cultural, strategic and personal narratives connect [CI15].

One of the main advantages of the proposed approach is that it aligns with existing social science models, making it easy for domain experts to understand and validate the annotations. This is in contrast to machine learning based approaches, which function as a black box which may produce classifications which are hard to explain. This is especially relevant for radical text detection where machine learning approaches may be subject to discriminatory bias against minorities.

Our immediate next steps include refining our rule-based annotator (e.g. adding more rules to find occurrences of narratives we are currently missing) and adapting it to other types of radicalising texts such as on-line forums for right- and left-wing extremism and incel culture.

Acknowledgements Work supported by the European Commission under grant 740934 – TRIVALENT – and grant 770302 – Co-Inform – as part of the Horizon 2020 research and innovation programme.



References

- [Bet08] David Betz. The virtual dimension of contemporary insurgency and counterinsurgency. *Small Wars & Insurgencies*, 19(4):510–540, dec 2008.
- [CI15] Álvaro Carrera and Carlos A. Iglesias. A systematic review of argumentation techniques for multi-agent systems research. *Artificial Intelligence Review*, 44(4):509–535, dec 2015.
- [Com18] European Commission. A multi-dimensional approach to disinformation. Technical report, European Commission, 2018.
- [Du 17] Cind Du Bois. Literature Review on Radicalisation. Technical Report October, TRIVALENT, 2017.
- [FAA18] Miriam Fernandez, Moizzah Asif, and Harith Alani. Understanding the Roots of Radicalisation on Twitter. In *Proceedings of the 10th ACM Conference on Web Science - WebSci '18*, pages 1–10. ACM, 2018.
- [GRBM16] Daveed Gartenstein-Ross, Nathaniel Barr, and Bridget Moreng. The Islamic State’s Global Propaganda Strategy. *Terrorism and Counter-Terrorism Studies*, (March), 2016.
- [HGC11] Jeffrey R. Halverson, Jr. Goodall H. Lloyd, and Steven R. Corman. *Master narratives of Islamist extremism*. Palgrave Macmillan, 2011.
- [Kot18] Ioannis E Kotoulas. Ideological Principles of Jabhat al-Nusra in Al-Risalah Magazine. (December), 2018.
- [KT11] Michael King and Donald M. Taylor. The Radicalization of Homegrown Jihadists: A Review of Theoretical Models and Social Psychological Evidence. *Terrorism and Political Violence*, 23(4):602–622, sep 2011.
- [LCGPC17] Raúl Lara-Cabrera, Antonio Gonzalez-Pardo, and David Camacho. Statistical analysis of risk assessment factors and metrics to evaluate radicalisation in Twitter. *Future Generation Computer Systems*, 2017.
- [LSF⁺13] Stephan Lewandowsky, Werner G K Stritzke, Alexandra M Freund, Klaus Oberauer, and Joachim I Krueger. Misinformation, Disinformation, and Violent Conflict. *American Psychologist*, 68(7):487–501, 2013.
- [Rie05] Catherine K Riessman. Narrative Analysis. In *Narrative, Memory & Everyday Life*, pages 1–7. 2005.
- [SDK⁺17] Hassan Saif, Thomas Dickinson, Leon Kastler, Miriam Fernandez, and Harith Alani. A Semantic Graph-Based Approach for Radicalisation Detection on Social Media. In *ESWC*, pages 571–587. Springer, 2017.