

Towards the Automatic Analysis of the Structure of News Stories

Iqra Zahid¹, Hao Zhang², Frank Boons³ and Riza Batista-Navarro²

¹School of Arts, Languages and Cultures

²School of Computer Science,

³Alliance Manchester Business School

Narratives

Analysis requires understanding of sequence of events

Most types follow chronological order (novels, personal accounts)

News stories

- follow *complex time structure*
- writers prioritise certain *news values*
 - negativity
 - unexpectedness
 - superlativeness

News stories

typically written using the
instalment method

- an event introduced earlier is described in detail (revisited) later
- possibly in multiple, separate instances

events tend to be presented in
non-chronological order

US troops ambushed in Honduras

TEGUCIGALPA

UNITED STATES troops in Honduras were put on high alert after at least six American soldiers were wounded, two seriously, in a suspected leftist guerrilla ambush yesterday, United States officials said.

Six or seven soldiers were wounded when at least three men, believed to be leftist guerrillas, used high-powered weapons in an ambush of a bus carrying 28 passengers 20 kilometres north of the capital Tegucigalpa, United States embassy spokesman Terry Kneebone said.

The bus was carrying the soldiers from a pleasure trip at a beach on the Atlantic Coast.

"It was a surprise attack," Southern Command spokesman Captain Art Haubold said in Panama City.

"The US forces did not return fire. They kept going to get out of the area as quickly as possible."

A Tegucigalpa radio station said an unidentified caller said the leftist group Morazanista Patriotic Liberation Front claimed responsibility for the attack. — NZPA-Reuter

The Dominion, Wellington, New Zealand, 2 April 1990

News schema

helps in understanding the flow of events

overall form of news discourse by which topics are organised

a set of formal categories that form the basis of organisation of textual units

seminal work by van Dijk (1985)

each category corresponds to a piece of text, i.e., a sequence of sentences

News narrative analysis

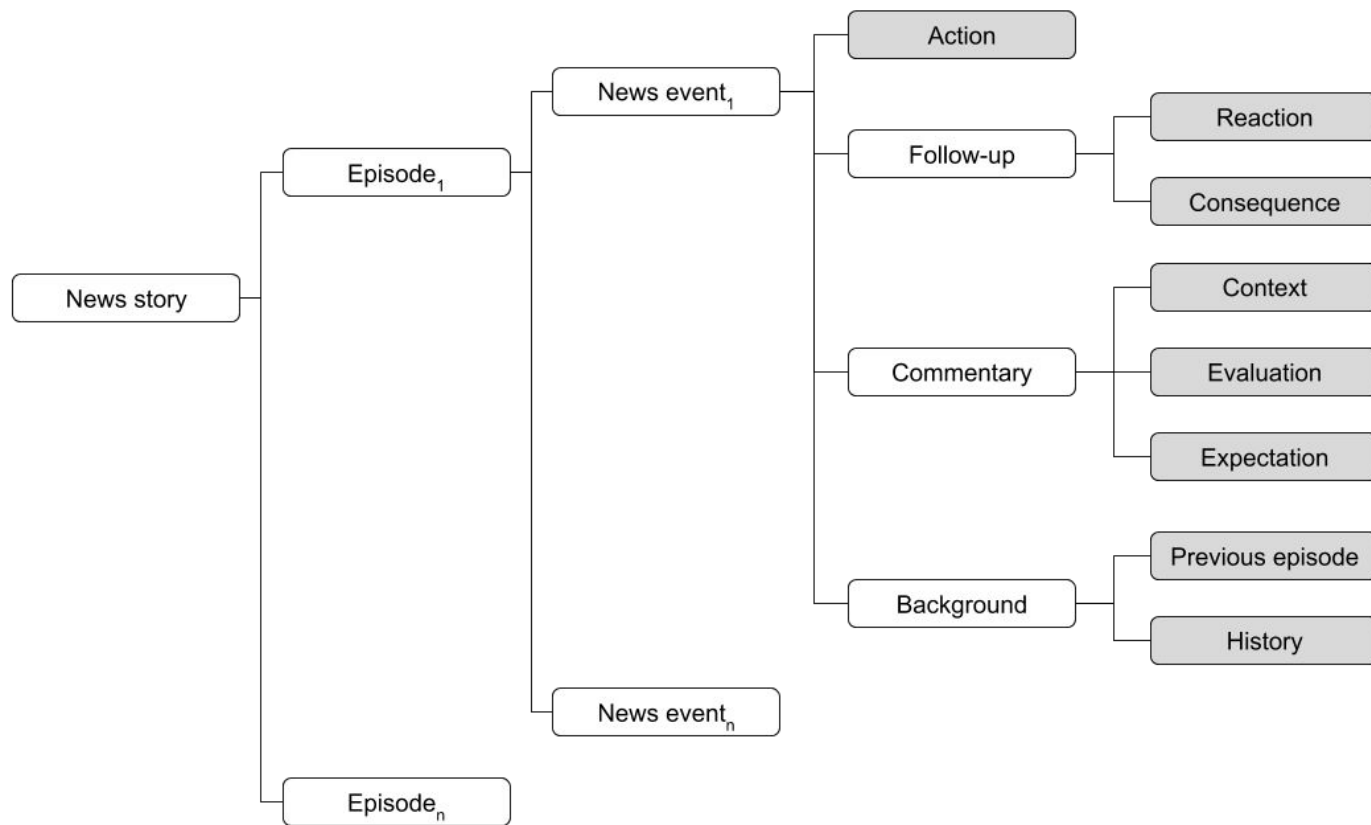
More challenging to automate

Event timelines difficult to reconstruct

Aim:

to facilitate machine understanding of news stories by automatically decomposing them according to news schema categories

News schema by Allan Bell



News schema by Allan Bell

Category	Definition	Example
Action	Central or main action	<i>More than a dozen students from Peking University in Beijing, in a rare rebuke of authority, protested Friday on campus to draw attention to the university's attempts to punish students for taking part in the campaign.</i>
Reaction	A response to the Action that was expressed verbally, e.g., a direct/indirect quote, speech	<i>Peking University officials moved swiftly to contain Friday's protest, holding the students in classrooms and keeping them through the night for questioning, activists said.</i>
Consequence	An action that transpired as a result of the Action, excluding verbal responses	<i>The students have put the government in an awkward position because they are invoking the teachings of Mao, Marx and Lenin, which President Xi Jinping has championed, to point to problems in Chinese society including inequality, corruption and greed.</i>

News schema by Allan Bell

Category	Definition	Example
Context	Additional information that help explain or clarify details surrounding the Action	<i>The students are part of a small but tenacious group of young communists using leftist ideology to shine a light on labor abuses across China and to call for better protections for the working class.</i>
Evaluation	An assessment of the significance of the Action	<i>The stern reaction by the authorities reflects the party's deep anxieties about the young communists and their unusual campaign.</i>
Expectation	A view on what could happen after the Action	<i>Party leaders may be concerned that the 30th anniversary of the massacre, coming up in June, could inspire new protests.</i>

News schema by Allan Bell

Category	Definition	Example
Previous episode	Events that happened more recently (in the near past)	<i>The protest on Friday came after Peking University officials tried to block a Marxist student group from organizing a celebration for Mao's 125th birthday.</i>
History	Set of events that happened before the near past	<i>The party has long feared student-led protests, especially since the 1989 pro-democracy movement, which had deep student involvement and was crushed in a bloody crackdown around Tiananmen Square.</i>

Corpus development

22 news articles collected using LexisNexis

- pre-2005 set
- post-2005 set

Annotation guidelines

- categories assigned to individual sentences
- only one of the most specific news schema categories assigned
- sentences that pertain to Action are identified first

Corpus development

Two annotators

Using *brat*

IAA (post-2005 set):

0.77 Kappa

The screenshot displays the *brat* web-based annotation interface. The main window shows a list of text segments, each with one or more annotations highlighted in green boxes. The annotations are labeled with terms like 'Consequence', 'Reaction', 'Action', 'Evaluation', 'History', and 'Expectation'. On the right side, an 'Edit Annotation' dialog box is open, showing the selected text: 'Party leaders may be concerned that the 30th anniversary of the massacre, coming up in June, could inspire new protests.' Below the text, there is a 'Search' field containing 'Google, Wikipedia'. The 'Entity type' section shows a list of categories: 'Consequence', 'Reaction', 'Commentary', 'Context', 'Evaluation', 'Expectation' (which is selected), and 'Background'. At the bottom of the dialog, there are buttons for 'Add Frag.', 'Delete', 'Move', 'OK', and 'Cancel'.

Linguistic devices

	Observation	Method
Action	shared lexical similarities with news title	exact matches between the lemmatised words of a sentence and title
Reaction	use of attribution, i.e., “who expressed what”; mentions that co-refer to either the Action itself or actors participating in it	attribution extraction (Zhang, 2019); detection of co-referring mentions (definite NPs)
Consequence	mentions that co-refer to either the Action itself or actors participating in it; discourse connectives signifying causation (e.g., “as a result”, “because”, “thereby”)	detection of co-referring mentions (definite NPs); use of discourse connectives in the Penn Discourse Treebank denoting the Contingency relation (min frequency = 4)

Linguistic devices

	Observation	Method
Evaluation	use of attribution; use of adjectives indicating assessment, i.e., the degree to which a quality holds (e.g., “deep”, “strongest”, “biggest”)	attribution extraction (Zhang, 2019); 260 graded adjectives collected from the Collins Cobuild Grammar Patterns
Expectation	use of attribution; use of speculative language	attribution extraction (Zhang, 2019); presence of modal verbs (e.g., “could”, “may”) and modifiers that indicate uncertainty from the WikiWeasel 2.0 corpus
Context	use of attribution; prevalent use of co-referring mentions	attribution extraction (Zhang, 2019); detection of co-referring mentions (definite NPs) and whether they appear as either the subject or object of sentences

Linguistic devices

	Observation	Method
Previous episode	main verbs in the past or past perfect tense; use of relative temporal expressions pertaining to recent points in time (e.g., “last week”, “previously”, “on Friday”)	rules and regular expressions
History	main verbs in the past or past perfect tense; use of absolute temporal expressions (e.g., “in 1989”); sometimes, use of relative temporal expressions pertaining to points in time in the distant past (e.g., “three decades ago”)	rules and regular expressions

Preliminary results

NLP Pipeline

- LingPipe Sentence Splitter
- GENIA Tagger (Tokenisation)
- Enju Parser (POS Tagging and dependency parsing)

Rules implemented in Python

Overall performance: F-score of 64% (precision = 70%, recall = 59%)

Ongoing and future work

Expanding the corpus

Observe any further linguistic devices

Extend our rules

Explore machine learning-based models

Make the corpus publicly available

Thank you!

Any questions?