Discourse Cues for Broadcast News Segmentation

Mark T. Maybury
The MITRE Corporation
202 Burlington Road
Bedford, MA 01730, USA
maybury@mitre.org

Abstract

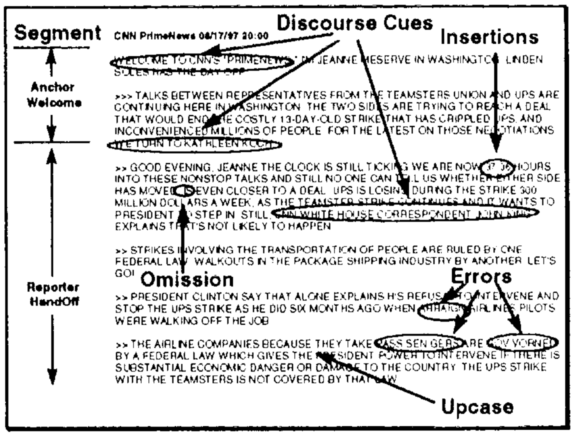
This paper describes the design and application of time-enhanced, finite state models of discourse cues to the automated segmentation of broadcast news. We describe our analysis of a broadcast news corpus, the design of a discourse cue based story segmentor that builds upon information extraction techniques, and finally its computational implementation and evaluation in the Broadcast News Navigator (BNN) to support video news browsing, retrieval, and summarization.

1. Introduction

Large video collections require content-based information browsing, retrieval, extraction, and summarization to ensure their value for tasks such as real-time profiling and retrospective search. Whereas image processing for video indexing currently provides low level indeces such as visual transitions and shot classification (Zhang et al. 1994), some research has investigated the use of linguistic streams (e.g., closed captions, transcripts) to provide keyword-based indexes to video. Story-based segmentation remains illusive. For example, traditional text tiling approaches often undersegment broadcast news because of rapid topic shifts (Mani et al. 1997). This paper takes a corpus-based approach to this problem, building linguistic models based on an analysis of a digital collection of broadcast news, exploiting the regularity utilized by humans in signaling topic shifts to detect story segments.

2. Broadcast News Analysis

Human communication is characterized by distinct discourse structure (Grosz and Sidner 1986) which is used for a variety of purposes including managing interaction between participants, mitigating limited attention, and signaling topic shifts. In processing genre such as technical or journalistic texts, programs can take advantage of explicit discourse cues (e.g., "the first", "the most important") to perform tasks such as summarization (Paice 1981). Our initial inability to segment topics in closed caption news text using thesaurus based subject assessments (Liddy and Myaeng 1992) motivated an investigation of explicit turn taking signals (e.g., anchor to reporter handoff). We analyzed programs (e.g., CNN PrimeNews) from an over one year corpus of closed caption texts with the intention of creating models of discourse and other cues for segmentation.



CNN PrimeNews 08/17/97 20:00

WELCOME TO CNN'S "PRIMENEWS." I'M JEANNE MESERVE IN WASHINGTON LINDEN SOLES HAS THE DAY OFF

>>> TALKS BETWEEN REPRESENTATIVES FROM THE TEAMSTERS UNION AND UPS ARE CONTINUING HERE IN WASHINGTON THE TWO SIDES ARE TRYING TO REACH A DEAL THAT WOULD END THE COSTLY 13-DAY-OLD STRIKE THAT HAS CRIPPLED UPS AND INCONVENIENCED MILLIONS OF PEOPLE FOR THE LATEST ON THOSE NEGOTIATIONS WE TURN TO KATHLEEN KOCH

>> GOOD EVENING, JEANNE THE CLOCK IS STILL TICKING WE ARE NOW 37 36 HOURS INTO THESE NONSTOP TALKS AND STILL NO ONE CAN TELL US WHETHER EITHER SIDE HAS MOVED IS EVEN CLOSER TO A DEAL UPS IS LOSING DURING THE STRIKE 300 MILLION DOLLARS A WEEK, AS THE TEAMSTER STRIKE CONTINUES AND IT WANTS TO PRESIDENT TO STEP IN STILL CNN WHITE HOUSE CORRESPONDENT JOHN KING EXPLAINS THAT'S NOT LIKELY TO HAPPEN

>> STRIKES INVOLVING THE TRANSPORTATION OF PEOPLE ARE RULED BY ONE FEDERAL LAW WALKOUTS IN THE PACKAGE SHIPPING INDUSTRY BY ANOTHER LET'S GO!

>> PRESIDENT CLINTON SAY THAT ALONE EXPLAINS H'S REFUSAL TO INTERVENE AND STOP THE UPS STRIKE AS HE DID SIX MONTHS AGO WHEN ARRAIGN AIRLINES PILOTS WERE WALKING OFF THE JOB

>> THE AIRLINE COMPANIES BECAUSE THEY TAKE PASS SEN GERS ARE GOV VORNED BY A FEDERAL LAW WHICH GIVES THE PRESIDENT POWER TO INTERVENE IF THERE IS SUBSTANTIAL ECONOMIC DANGER OR DAMAGE TO THE COUNTRY THE UPS STRIKE WITH THE TEAMSTERS IS NOT COVERED BY THAT LAW

Figure 1. Closed Caption Challenges
(CNN Prime News, August 17, 1997)

While human captioners employ standard cues to signal discourse shifts in the closed caption stream (e.g., ">>" is used to signal a speaker shift whereas ">>>" signals a subject change), these can be erroneous, incomplete, or inconsistent. Figure 1 illustrates a typical excerpt from our corpus. Our creation of a gold standard corpus of a variety of broadcast sources indicates that transcription word error rates range from 2% for pre-recorded programs such as 60 Minutes news magazine to 20% for live transcriptions (including errors of insertion, deletion, and transposition). This noisy data complicates robust story segmentation.

2.1 News Story Discourse Structure

Broadcast news has a prevalent structure with often explicit cues to signal story shifts. For example, analysis of the structure of ABC World News Tonight indicates:

- broadcasts start and end with the anchor
- reporter segments are preceded by an introductory anchor segment and together they form a single story
- commercials serve as story boundaries

Similar but unique structure is also prevalent in many other news programs such as CNN Prime News (See Figure 1) or MS-NBC. For example, the structure for the Jim Lehrer News Hour provides not only segmentation information but also *content* information for each segment. Thus, the order of stories is consistently:

- preview of major stories of the day or in the broadcast program
- sponsor messages
- summary of the day's news (including some major stories)
- four to six major stories
- recap summary of the day's news
- sponsor messages

Recovering this structure would enable a user to view the four minute opening summary, retrieve daily news summaries, preview and retrieve major stories, or browse a video table of contents, with or without commercials.

2.2 Discourse Cues and Named Entities

Manual and semi-automated analysis of our news corpora reveals that regular cues are used to signal these shifts in discourse, although this structure varies dramatically from source to source. For example, CNN discourse cues can be classified into the following categories (examples from 8/18/97):

- Start of Broadcast
  `"GOOD EVENING, I'M KATHLEEN KENNEDY, SITTING IN FOR JOIE CHEN."`
- Anchor-to-Reporter Handoff
  `"WE'RE JOINED BY CNN'S CHARLES ZEWE IN NEW ORLEANS. CHARLES?`
- Reporter-to-Anchor Handoff
  `"CHARLES ZEWE, CNN, NEW ORLEANS"`
- Cataphoric Segment
  `"STILL AHEAD ON PRIMENEWS"`
- Broadcast End
  `"THAT WRAPS UP THIS MONDAY EDITION OF "PRIMENEWS""`

The regularity of these discourse cues from broadcast to broadcast provides an effective foundation for discourse-based segmentation routines. We have similarly discovered regular discourse cues in other news programs. For example, anchor/reporter and reporter/anchor handoffs in CNN Prime News or ABC News and other network programs are identified through pattern matching of strings such as:

- (word) (word) ", ABC NEWS"
- "ABC'S CORRESPONDENT" (word) (word)

The pairs of words in parentheses correspond to the reporter's first and last names. Combining the handoffs with structural cues, such as knowing that the first and last speaker in the program will be the anchor, allow us differentiate anchor segments from reporter segments. By preprocessing the closed caption text with a part of speech tagger and named entity detector (Aberdeen et al. 1995) retrained on closed captions, we generalize search of text strings to the following class of patterns:

- (proper name) ", ABC NEWS"
- "ABC'S CORRESPONDENT" (proper name)

3. Computational Implementation

Our discourse cue story segmentor has been implemented in the context of a multimedia (closed captioned text, audio, video) analysis system for web based broadcast news navigation. We employ a finite state machine to represent discourse states such as an anchor, reporter, or advertisting segment (See Figure 2). We further enhance these with multimedia cues (e.g. detected silence, black or logo keyframes) and temporal knowledge (indicated as time in Figure 2). For example, from statistical analysis of CNN Prime News Programs, we know that weather segments appear on average 18 minutes after the start of the news.

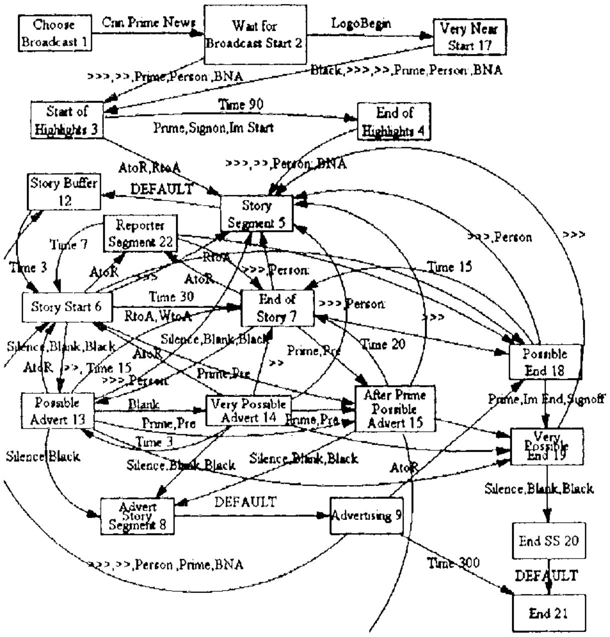


Figure 2. Partial Time-Enhanced FSM

After segmentation, the user is presented with a hierarchical navigation space of the news which enables search and retrieval of segmented stories or browsing stories by date, topic, named entity or keyword (see Figure 3). This is MITRE's Broadcast News Navigator (http://www.mitre.org/resources/centers/advanced_info/g04f/bnn/mmhomeext.html).

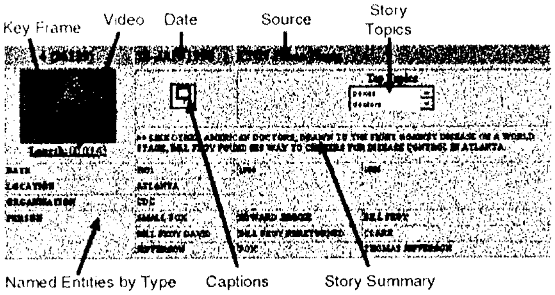


Figure 3. Broadcast News Navigator

We leverage the story segments and extracted named entities to select the sentence with the most named entities to serve as a single sentence summary of a given segment. Story structure is also useful for multimedia summarization. For example, we can select key frames or key words from the substructure which will likely contain the most meaningful content (e.g., an reporter segment within an anchor segment).

4. Evaluation

We evaluated segmentor performance by measuring both the precision and recall of segment boundaries compared to manual annotation of story boundaries where:

1. Precision $= \dfrac{\text{\# of correct segment tags}}{\text{\# of total segment tags}}$
2. Recall $= \dfrac{\text{\# of correct segment tags}}{\text{\# of hand tags}}$

Source	Precision	Recall
ABC World News	90	94
CNN Prime News	95	75
Jim Lehrer News Hour	77	52

Table 1. Segmentation Performance

Table 1 presents average precision and recall results for multiple programs after applying generalized cue patterns developed first for ABC as described in Section 2.2. Recall degrades when porting these same algorithms to different news programs (e.g., CNN, Jim Lehrer) given the genre differences as described in Section 2.1.

Errors in story boundary detection include erroneously splitting a single story segment into two story segments, and merging two contiguous story segments into a single story segment. Furthermore, given our error-driven transformation based proper name taggers operate at approximately 80% precision and recall, this can adversely impact discourse cue detections. Also, our preliminary evaluation of speech transcription results in word error rates of approximately 50%, which suggest non captioned text is not yet feasible for this class of segmentation.

We have just completed an empirical study (Merlino and Maybury, forthcoming) with BNN users that explores the optimal mixture of media elements show in Figure 3 (e.g., keyframes, named entities, topics) in terms of speed and accuracy of story identification and comprehension tasks. Key findings include that users perform better and prefer mixed media presentations over just one media (e.g., named entities or topic lists), and they are quicker and more accurate working from extracts and summaries than from the source transcript or video.

6. Conclusion and Future Work

We have described and evaluated a news story segmentation algorithm that detects news discourse structure using discourse cues that exploit fixed expressions and transformational-based, part of speech and named entity taggers created using error-driven learning. The implementation utilizes a time-enhanced finite state automata that represents discourse states and their expected temporal occurance in a news broadcast based on statistical analysis of the corpus. This provides an important mechanism to enable topic tracking, indeed we take the text from each segment an run this through a commercial topic identification routine an provide the user with a list of the top classes associated with each story (See Figure 3). The segmentor has been integrated into a system (BNN) for content-based news access and has been deployed in a corporate intranet and is currently being evaluated for deployment in the US government and a national broadcasting corporation.

We have improved segmentation performance by exploiting cues in audio and visual streams (e.g., speaker shifts, scene changes) (Maybury et al. 1997). To obtain a better indication of annotator reliability and for comparative evaluation, we need to measure interannotator agreement. Future research includes investigating the relationship of other linguistic properties, such as co-reference, intonation contours, and lexical semantics coherence to serve as a measure of cohesion that might further support story segmentation. Finally, we are currently evaluating in user studies which mix of media elements (e.g., key frame, named entities, key sentence) are most effective in presenting story segments for different information seeking tasks (e.g., story identification, comprehension, correlation).

Acknowledgements

Andy Merlino is the principal system developer of BNN. The Alembic sub-system is the result of efforts by MITRE's Language Processing Group including Marc Vilain and John Aberdeen for part of speech proper name taggers, and David Day for training these on closed caption text.

References

Aberdeen, J.; Burger, J.; Day, D.; Hirschman, L.; Robinson, P. and Vilain, M. (1995) “Description of the Alembic System Used for MUC-6”, Proceedings of the Sixth Message Understanding Conference, Columbia, MD, 6-8 November, 1995.

Brill, E. (1995) Transformation-based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging. *Computational Linguistics*, 21(4).

Grosz, B. J. and Sidner, C. July-September, (1986) "Attention, Intentions, and the Structure of Discourse." *Computational Linguistics* 12(3):175-204.

Liddy, E. and Myaeng, S. (1992) "DR-LINK's Linguistic-Conceptual Approach to Document Detection", Proceedings of the First Text Retrieval Conference, 1992, NIST.

Mani, I., House, D., Maybury, M. and Green, M. (1997) Towards Content-based Browsing of Broadcast News Video. In Maybury, M. (ed.) *Intelligent Multimedia Information Retrieval*, AAAI/MIT Press, 241-258.

Merlino, A. and Maybury, M. forthcoming. An Empirical Study of the Optimal Presentation of Multimedia Summaries of Broadcast News. In Mani, I. and Maybury, M. (eds.) *Automated Text Summarization*

Merlino, A., Morey, D. and Maybury, M. (1997) "Broadcast News Navigation using Story Segments", Proceedings of the ACM International Multimedia Conference, Seattle, WA, November 8-14, 381-391.

Paice, C. D. (1981) The Automatic Generation of Literature Abstracts: An Approach Based on the Identification of Self-Indicating Phrases. In Oddy, R. N., Robertson, S. E. , van Rijsbergen, C. J., Williams, P. W. (eds.) *Information Retrieval Research.* London: Butterworths, 172-191.

Zhang, H. J.; Low, C. Y.; Smoliar, S. W. and Zhong, D. (1995) Video Parsing, Retrieval, and Browsing: An Integrated and Content-Based Solution. Proceedings of ACM Multimedia 95. San Francisco, CA, p. 15-24.