

Khalid Al-Khatib, Henning Wachsmuth, Johannes Kiesel, Matthias Hagen, Benno Stein

A News Editorial Corpus for Mining Argumentation Strategies

Argumentation Strategies

In editorials, authors follow a specific strategy to persuade their readers of some opinion or attitude. This includes:

- (1) Select a set of argumentative units.
- (2) Arrange them.



A new annotation scheme for mining argumentation strategies that captures the type of each argumentative discourse unit of an argumentative text.

Title. An education was my path to financial security. Then I got my student loan bill.

Editorial. I have a very distinct memory from my first day of college. My family's minivan slowly pulling into my dormitory's parking lot, through a crowd of first-year students flanked by helicopter parents and, in retrospect, probably hungover orientation week advisers. I remember thinking "Hurry up! I'm ready to start my real life!"

I had no idea what I was really pushing towards. As the only daughter of Nigerian immigrants with a tenuous-at-best toehold on the middle class, college was billed as the only path to financial security. "No one can ever take away your education," my father would say repeatedly. While that may be true, two degrees later someone could take away my access to decent housing because of my shit credit, thanks to the nearly \$60,000 in student loans I've essentially defaulted on since graduating from the University of Chicago and Northwestern University.

It seems a college education is part of the American dream that's easy to buy (or borrow) into, but hard to pay off. With tuition soaring, and the middle class shrinking along with their incomes, many students and their families are left holding incredibly expensive bags. In 2013, 69% of graduating seniors at public and private nonprofit colleges took out student loans to pay for college, and "about one-fifth of new graduates' debt was in private loans," according to the Project on Student Debt. Even public schools - long considered a more affordable option - are less accessible: public colleges increasingly rely on tuition dollars as state funding continues to fall (25% and 23%, respectively, in 2012, compared to 17% and 23% in 2003). The country's cumulative student loan debt (\$1.1m) has surpassed car loans (\$875bn) and credit card debt (\$659bn). Though college graduates make more than their peers who only graduated from high school, for many, monthly student loans leech into that extra \$17,500 in salary.

Yet the party line that college education is the middle class' only hope for upward mobility persists - it will even be the message of President Obama's last stop on his "SOTU Spouter" tour in Knoxville, Tennessee.

"In today's economy," Dan Pfeiffer, the president's senior advisor, wrote on Medium, "access to a college education is the surest ticket to the middle class -- and the President's proposals will help more young people punch that ticket."

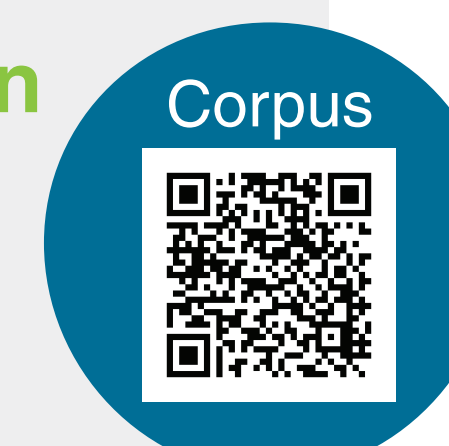
Is someone who punched that ticket twice. I'm still waiting for my express bus to the middle class. The modest income I make as an entrepreneur with a day job is whittled away each month thanks to loan payments (plus interest) to various financial institutions that feel more like bounty hunters than supporters of middle-class aspirations.

With that \$60,000 in student loans hanging over me, I'm still waiting to start the "real" life I'd always imagined for myself. It's just that now I want one with its possibilities a little less hampered by student debt.

- Types of units
- Anecdote
 - Assumption
 - Common ground
 - Other
 - Statistics
 - Testimony

Main contributions of our work:

- The first corpus which allows studying **argumentation strategies** in monological opinionated text.
- Detailed insights into the **corpus construction**.
- **Corpus analysis** in terms of the agreement and distribution of types within and across news portals.



Corpus Construction

300 editorials are selected randomly and evenly from Al Jazeera, Fox news, and Guardian news portals.

Type	Editorials	Total	Std. dev.	Median	Min	Max
Tokens	All editorials	287364	257.28	932	298	1894
	Al Jazeera	106430	236.05	1033	440	1671
	Fox News	86415	226.36	855	298	1613
Sentences	All editorials	11754	13.00	37	12	114
	Al Jazeera	3962	10.55	38	16	75
	Fox News	3912	13.45	39	12	104
	Guardian	3880	14.65	36	18	114

Editorials are segmented semi-automatically into argumentative discourse units using a novel segmentation algorithm.

Type	Editorials	Total	Std. dev.	Median	Min	Max
Segments	All editorials	35665	38.21	116	28	309
	Al Jazeera	11521	31.68	113	32	218
	Fox News	11315	35.4	112	28	231
	Guardian	12829	44.58	122	59	309



Each argumentative discourse unit is labeled manually with one of six types.

Type	Total	Std. dev.	Median	Min	Max	Percent
Common ground	241	1.53	0	0	13	1.7%
Assumption	9792	12.42	32	3	86	68.4%
Anecdote	2603	8.68	7	0	77	18.2%
Testimony	1089	5.42	2	0	44	7.6%
Statistics	421	2.76	0	0	19	2.9%
Other	167	1.64	0	0	24	1.2%
All units	14313	14.28	46	14	132	100%

High inter-annotator agreement in terms of Fleiss' k is obtained considering the number and complexity of the labels.

	Comm.	Ass.	Anec.	Test.	Statistics	Other	Overall
Fleiss' k	0.114	0.613	0.399	0.591	0.582	0.152	0.560

Corpus Analysis

Disagreement analysis based on the confusion probability matrix (CPM, Cinková et al. (2012)).

	Common gro.	Assumption	Testimony	Statistics	Anecdote	Other
Common gro.	0.129	0.562	0.012	0.005	0.163	0.012
Assumption	0.035	0.701	0.017	0.01	0.075	0.014
Testimony	0.006	0.134	0.618	0.016	0.087	0.002
Statistics	0.006	0.195	0.042	0.603	0.074	0.002
Anecdote	0.037	0.277	0.041	0.013	0.451	0.016
Other	0.018	0.324	0.006	0.003	0.101	0.166

Distribution of types of argumentative discourse units among the three news portals.

