**Annotation Guideline**

**Objective:**

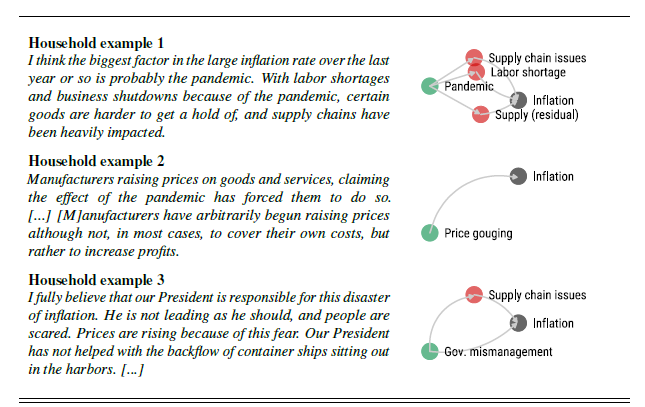
The objectives of this study are to analyze **media narratives discussing the causes of inflation** and observe changes over time

**What is a narrative?**

According to Shiller (2017), **narratives are understood as contagious and popular stories** that spread virally and are seen as a central cause of economic fluctuations.

Abbot (2002) highlights the importance of events. For him, **narratives are always a representation of an event or multiple events**, and thus from a literature study perspective distinguishable from a description, lyric or argument.

In line with this idea, Eliaz and Spiegler (2020) define a **narrative as a causal model** represented by a 'Directed Acyclic Graph' (DAG). The model itself can therefore vary in terms of nodes and relations.

Closely related to our work is the study by Andre et al. (2023). They studied inflation narratives among households and experts in the US by means of open-ended questions and created **DAG representations** by hand coding the answers. Each node represents an event, while the edges represent relations between them:

**Narratives vs. other text types**

Narrative can be fictional or real, exciting or dry. Common among narratives is the fact that they describe the **experiences of one or more entities** involved in a **sequence of events**. Narratives are typically written in the past tense, but can also appear in the present tense.

1. **Entities/Character**: People (Characters), Institutions, Animals etc.
2. **Events**: Generally understood as “occurrences at a particular place and time”, involve verbs, but can be also nouns/adjectives

These two elements are crucial for the (linguistic) definition of a narrative. As Abbot (2002) points out, these two elements make up the **narrative’s story**. However, besides the story there is also the **narrative discourse**. This discourse mediates the story (by voice, style etc.), but the story itself is never defined by these discursive elements. Instead “the story lies in the events, situations, and behaviors signified by the words” (image etc.) (reference: Bremond)

In our pretest study, we try to identify narratives by **focussing on the story element**, which is the core feature that helps to distinguish it from other forms of telling.

**Data:**

Documents presented to you are coming from the **Dow Jones Newswire**. The dataset includes content in the following areas: market-moving M&A, exclusives, and earnings news; full-text feeds from Dow Jones sources (Newswires, The Wall Street Journal, Barron’s, MarketWatch); global company news; central bank, macroeconomic, political, FX, commodities and energy news; third-party press release wires (BusinessWire, PR Newswire, Globe Newswire and others).

The documents are pre-filtered based on a two-step procedure:

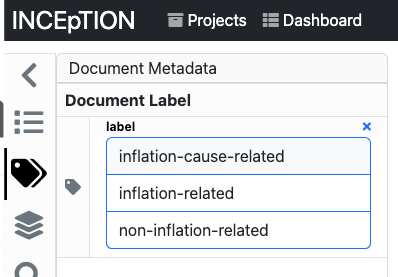
1. We applied a simple **keywords based filtering** including the expressions of inflation and prices
2. Further, by means of the Dow Jones Newswire **subject codes** we selected a number of relevant news sources. The selection of subject codes was guided by the interest of analyzing newspaper-like texts (see appendix table).

For the following annotation task we will **randomly select 100 documents** from this pre-filtered corpus. All documents will be annotated by three annotators, allowing us to measure and evaluate the consensus.

**Annotation Task:**

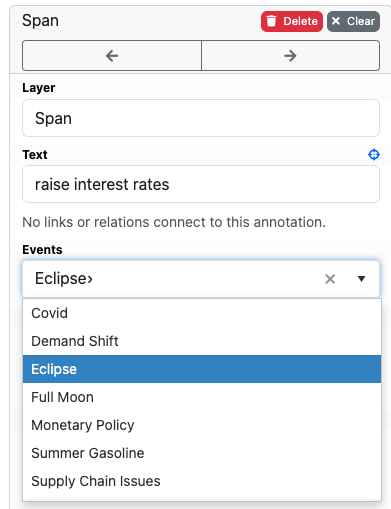
The overall objective of this work is to identify inflation narratives in large text corpora. Thus, the leading questions for the annotation task are:

1. **Does the present text discuss a (more multiple) cause(s) of inflation?**
   1. **inflation-cause-related: cause of inflation is mentioned in this narrative**
   2. **inflation-related: inflation is a part of the core story, but the core story is not limited to its cause, example topic can also be the effect of inflation etc.**
   3. **non-inflation-related: inflation is not a part of the core story**

****

This task involves distinguishing whether or not the present **text talks and argues about the causes of inflation**. This is a binary task, meaning there is only yes or no as an annotation tag.

1. **If so, which events are present in the text?**

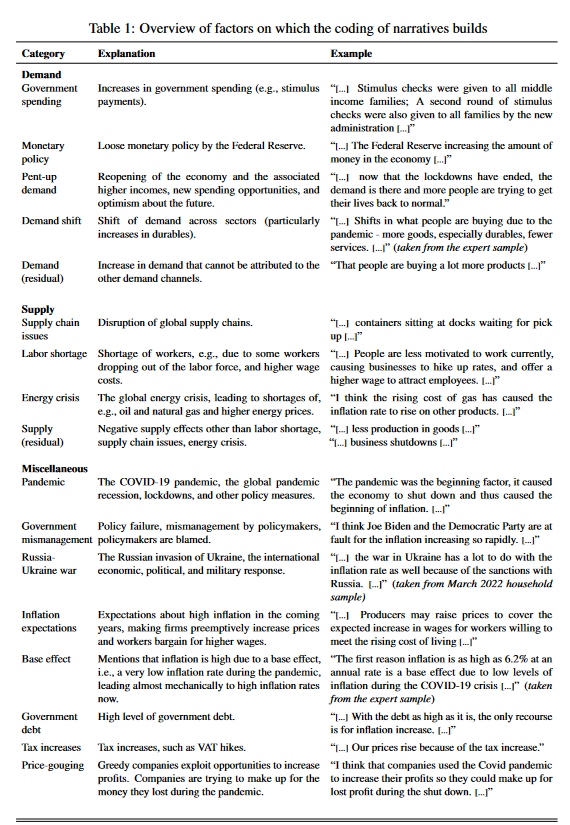
****

With this task we try to identify the narratives in the present text. As discussed earlier, a crucial part that distinguishes narratives from other text types is the fact that it includes a **sequence of events**, in which entities/characters play a role.

As a starting point we use the code system delivered by Andre et al. (2023) study on inflation narratives among households and experts. We add the code “inflation”, which includes events that describe a rise in prices. The central task is therefore to **assign the different codes to sequences** in the text, e.g.:

*“The Federal Reserve's response to the COVID-19 shock, marked by slashing interest rates and injecting massive liquidity, was intended to stabilize the economy but ultimately contributed to rising inflation. By keeping these policies in place too long, even as demand rebounded and supply chains remained disrupted, the Fed fueled excessive demand and thus price increases.”*

This example serves as orientation and is generated by ChatGPT4. The story of this paragraph is a collection of events involving the event of 1. Monetary Policy (yellow), 2. “Pent-up demand”, 3. “Supply chain issues”, 4. “Demand residuals”, and 5. “inflation”.



* “Inflation” - “Increase in prices” - “ … therefore prices increased”

1. **How are these events related?**

After event annotation in the present text, the next task involves the annotation of the **relations between these identified events**. A relation is therefore only of interest if it's regarding the **“causal structure”,** meaning if it’s about the cause of one event. Thus expressions like “due to, thus, and caused by” serve as indicators. However, sometimes the causal links are not very explicit, which is a challenge for annotators and LLMs.

In the example above, the relations between the elements could be understood as the following:

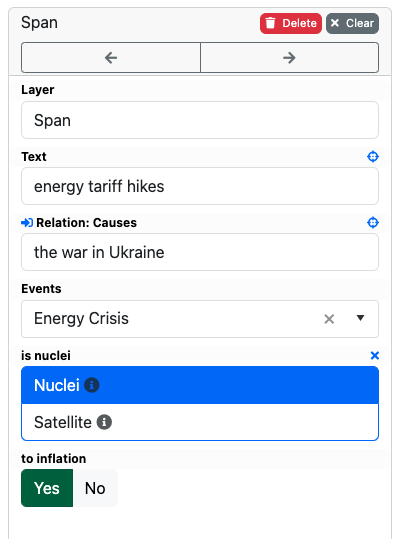
1. **What is the document’s core story (authorial intents)?**

The final annotation task involves identifying the core story of the document. This is challenging, because it includes **filtering out all side stories** that are present in the document.

**Constituent events:** Essential events, or nuclei (backbone of the narrative), without them, the story would fundamentally change.

**Supplementary events:** Non-essential events, or satellite (catalyst), that add depth, richness, and complexity to the story.

The core story is constructed with a sequence of constituent events. These constituent events do not refer to all the causal events to inflation. Rather, the narrator might mention K number of causes, but only N (N<K) causes construct the core story as an opinion/authorial intent. In the above example, constituent events will be: monetary policy, demand and inflation. Supplementary events are pent-up demand and supply chain issues.



According to Abbot (2002), if we would remove the supplementary events , the story would be less detailed or interesting, maybe also less convincing, but be still recognizable as the same story.

We are interested in filtering the constituent events, because this would allow us to (a) highlight the authorial intents and (b) reduce the sparsity of relations which is due to the typically high number of events in a document.

**Example**

*“The Federal Reserve's response to the COVID-19 shock, marked by slashing interest rates and injecting massive liquidity, was intended to stabilize the economy but ultimately contributed to rising inflation. By keeping these policies in place too long, even as demand rebounded and supply chains remained disrupted, the Fed fueled excessive demand and thus price increases.”*

In this small example all constitutional events are related to monetary policy issues. The author’s intention is to highlight the importance of the monetary policy for the rising inflation rates. Supplementary events such as “demand rebounded” or “supply chains remained disrupted” are non-essentials, because they purpose is to add richness and complexity to the story, however they do not change the overall story.

**More about the project:**

Github-Page: <https://github.com/semantic-systems/cdl-inflation-narrative>

Annotation-Page: <https://annotation.hitec.skynet.coypu.org/>

**Appendix**

**Considered News Sources**

| Subject Code | Description |
| --- | --- |
| DJIB | Dow Jones Investment Banker |
| DJG | Dow Jones Institutional News |
| GPRW | Dow Jones Global Press Release Wire |
| DJAN | Dow Jones Australian/New Zealand Report |
| AWSJ | Wall Street Journal Asia |
| WSJE | Wall Street Journal Europe |
| PREL | Press Release Wires |
| NRG | Dow Jones Energy Service |
| DJBN | Dow Jones Global News Select |
| AWP | AWP News |
| BRNS | Barron’s |
| JNL | Wall Street Journal - Online Versions of Print Articles |
| WAL | Wall Street Journal (domestic) stories filed direct to Newswires |
| WLS | Wall Street Journal (all) on Newswires |
| WSJ | Wall Street Journal - PB |