

ClaimPT Annotation Guidelines

Version 1.2.0

September 2025

Purificação Silvano^{1,3}

Raquel Sequeira²

Sara Nerea²

Inês Cantante¹

Diogo Folques²

Ricardo Campos^{2,3}

1 University of Porto, Porto, Portugal

2 University of Beira Interior, Covilhã, Portugal

3 INESC TEC, Porto, Portugal

Preface

Nowadays, misinformation has been gaining ground in the virtual space and media ecosystem. Fact-checking is therefore becoming an indispensable activity. For this reason, numerous researchers have been trying to understand how fact-checking can be automated, either through the use of techniques based on natural language processing and machine learning, or through the use of databases to automatically predict the veracity of statements. Thus, although there have been some efforts to automatically verify the veracity of information, the truth is that this process is still very much concentrated on the English language. Furthermore, for the automatic detection of disinformation, i.e., false news or statements, the existence of annotated datasets is crucial. The ClaimPT project aims to develop a proposal for the detection of claims and non-claims for European Portuguese, starting with the creation of an annotated dataset consisting of news items from the Portuguese news agency, LUSA. This document is intended to serve as an annotation scheme and instructions for annotating news articles. The article “NEWSCLAIMS: A New Benchmark for Claim Detection from News with Attribute Knowledge,”¹ which takes into account elements such as topic, stance, claim object, and claimer, served as the basis for the development of this annotation scheme. At the same time, the article “Toward automated fact-checking: Detecting check-worthy factual claims by ClaimBuster”² was a reference, especially with regard to the opposition between claim and non-claim, also presenting strategies for resolving conflicts between annotators. As for the classification and determination of sources, we used the article “Classification of news sources”³ by Aldo António Schmitz and, simultaneously, the “Journalism Manual”⁴ by Anabela Gradim.

Keywords: Fact-Checking, News Articles, Detection of Claims and Non-Claims, Verifiability, Disinformation.

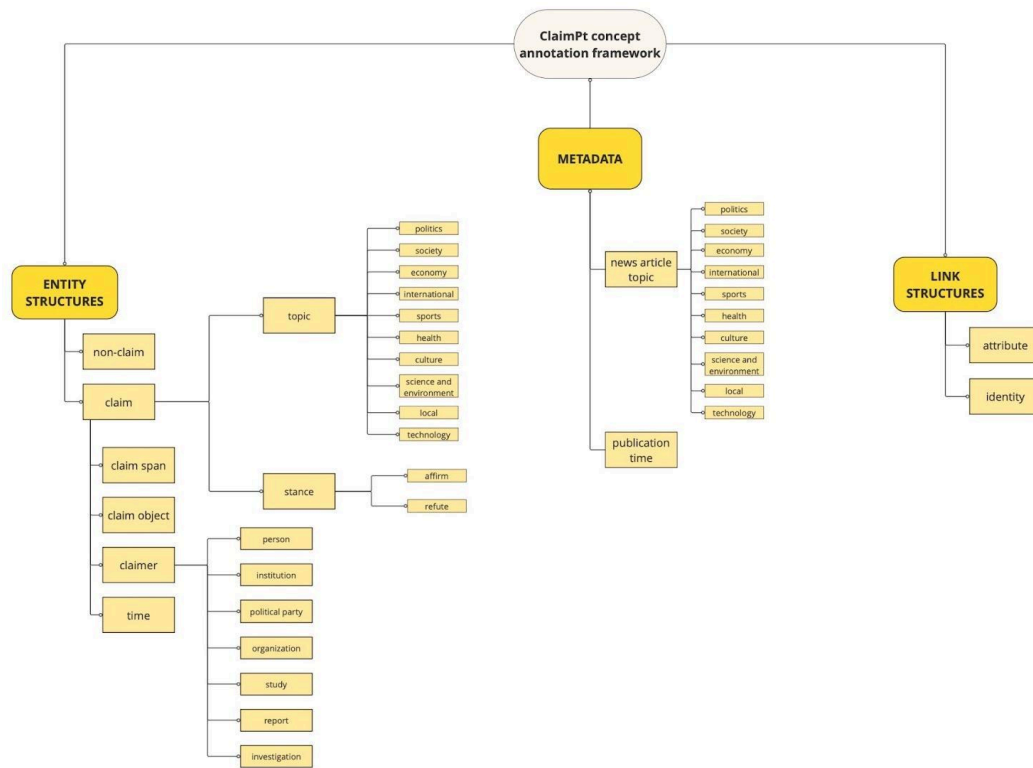
¹ Reddy, R. G., Chetan, S., Wang, Z., Fung, Y. R., Conger, K., Elsayed, A., ... & Ji, H. (2021). Newsclaims: A new benchmark for claim detection from news with attribute knowledge. arXiv preprint arXiv:2112.08544.

² Hassan, N., Li, C., Arslan, F., & Tremayne, M. (2017). Toward automated fact-checking: Detecting check-worthy factual claims by ClaimBuster. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1803-1812). ACM.

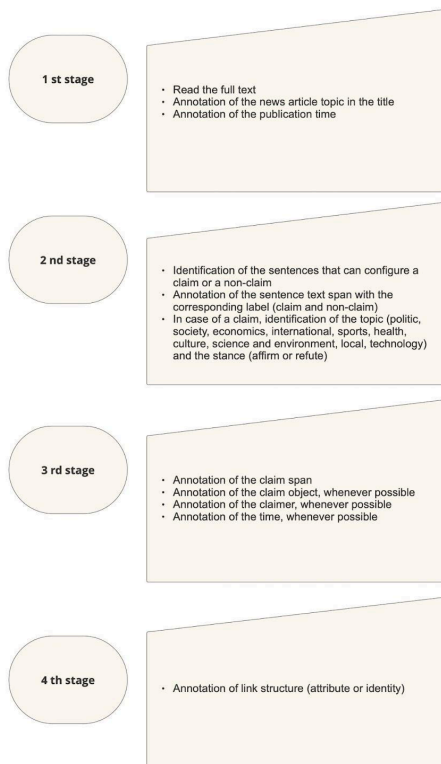
³ Schmitz, A. A. (2011). Classification of News Sources. Florianópolis: UFSC.

⁴ Gradim, A. (2000). Journalism Manual. Universidade da Beira Interior/Livros Labcom.

I. Annotation scheme structure



II. Stages of the annotation process



III. Annotation guidelines

Metadata

When annotating the topic, annotators should choose only one topic or class for each article, based on the title, using the label “News Article Topic.” The choice of topic should be the one that best fits, taking into account the subject matter of the piece. In section A.1, we list the topics that can be chosen in the articles. Additionally, the date on which the article was published must be annotated using the label “Publication Time.”

Definition of claim

A claim is a factual statement containing an alleged fact about the real world, of public interest, which can be verified. In the context of this project, only declarative sentences in direct speech, uttered exclusively by agents other than the journalist (such as people, associations, organizations) are identified as claims. In this sense, these sentences often appear associated with reporting verbs, such as “affirm,” “emphasize,” or “refer,” which precede or follow the agent responsible for the quote (the “claimer”).

The following list illustrates the notion of *claim* (identified in **bold**) as understood in this project.

1. “We began Internal Medicine month with the news that, unsurprisingly, **there were vacancies left to fill in the competition to choose specialties, and Internal Medicine was the hospital specialty with the most unfilled vacancies.**” (Text: Lusa 5)
2. “**This system worked until the spring of 2022 and worked well,** then it was changed by the European Commission and today truck drivers are blocking the border.” (Text: Lusa 12)

Entity structures

A. Claim

Definition: Factual statements that are of public interest. Generally, *claims* that contain information about health, politics, such as measures taken by the government, wars around the world, and climate change are of collective interest.

Text segment: Only text segments containing declarative sentences that occur in direct speech, with quotation marks and with complete meaning will be annotated. Quotation marks and full stops should not be included in the text segment.

- Each quotation must be analyzed individually (phrase by phrase).

In the example below, two *claims* are identified (one in **bold** and one in **blue bold**), as understood in this project.

Example: “**We are at the university station, which is the end of the line for the southern metro line of the Tagus, leaving thousands of people without alternative transportation. The government had promised to extend the light rail to Costa da Caparica.**” (Text: Lusa 60)

- Text segments that correspond to expressions or parts of incomplete sentences that do not contain a meaningful statement will not be annotated.

The following example illustrates an incomplete sentence (in **orange bold**).

Example: At the signing ceremony for the ‘Compromisso Visão Zero 2030’ the National Road Safety Authority and ‘Brisa Autoestradas’, part of the National Road Safety Strategy – Visão Zero 2030, José Luís Carneiro stated that the revenue from speed cameras, which is around 70 million to 80 million euros, “**will have been below forecasts.**” (Text: Lusa 13)

Example: The coordinator of the Bloco de Esquerda, Mariana Mortágua, said today that her party wants to “**heal the wounds of the socialist absolute majority**” look at the problems, and provide solutions. (Text: Lusa 60)

- Sentences that are factual, common sense, objective, or purely informative statements without opinion should not be labeled as *claims*.

In the example below, two claims are identified (one in **bold** and other in **blue bold**) and one purely informative sentence that cannot be annotated (in **purple**).

Example: “**We came from a right-wing country and a troika. It was this majority that allowed the country to look at itself with dignity and recover wages and pensions.**” (Text: Lusa 60)

Example: “**And from today, when the call opens, entities that are part of the Portuguese Contemporary Art Network will be able to apply for funding of two million euros,**” specified the minister of Culture. (Text: Lusa 7)

- There are phrases that could be *claims*, but as they are far removed from the reality in Portugal, they may be difficult for a journalist to verify quickly. Additionally, phrases referring to the future (or future possibilities) should not be annotated.

Example: “Approximately 25% of the minutes [excuses for overtime] will have an impact in December,” he explained. (Text: Lusa 5)

- The decision to annotate always rests with the annotator: if annotators encounter statements of public interest, they must decide whether or not it should be annotated. Note the example below, containing a claim about the Cumbie Fire and Rescue Service:

Example: “I am running out of water and having to ration it. **I know the emergency services are busy, but we are here without help.**” (Text: Lusa 226)

A.1. Topic

Definition: The topic corresponds to the theme underlying the claim (Reddy et al., 2021). Annotators should analyze only the claim in question, and not the entire text, in order to define the topic that best fits the claim, even when the scope (or general theme) of the article may have been classified with a different topic. If there is doubt between more than one topic, annotators should choose the one that is most salient in the *claim*.

The following topics are considered:

Politics - When claims related to the government, political parties, and national politics (e.g., election candidacies, approval of bills).

Example: “**The government had promised to extend the light rail to Costa da Caparica, and that promise was not kept,**” Mortágua criticized yesterday. (Text: Lusa 60)

Society - claims related to issues in Portuguese society such as justice, education, immigration, social protection, accidents (e.g., changes in tuition fees, high-profile cases in the Portuguese justice system, cases of police violence).

Example: **“Regarding the number of promotions and career advances in recent years and how many workers have not moved up in more than 3, 5, and 10 years, it is clear that most workers have not had any professional advancement in the last 10 years.”** (Text: Lusa 25)

Economy - All claims related to the national economy, monetary policy, banking, public finances, labor and employment issues (e.g., approval of the state budget, changes to tax law, increase in public debt).

Example: **“Officials come to boast about budgetary achievements, but the CGE reveals that the real achievement belongs to taxpayers, and more than half of the surplus results from investment that was never executed,”** said. (Text: Lusa 536)

International - Claims related to other countries, conflicts, and organizations (e.g., NATO meetings, armed conflicts, natural disasters).

Example: **“The Israeli machine of death, destruction, and genocide has resumed the murder of children, women, and the elderly, ignoring all reason, conscience, and humanity.”**(Text: Lusa 91)

Sports - claims about different sports and national athletes (e.g., refereeing issues, club management issues).

Example: **“There are still many drivers who have not signed a contract.”** (Text: Lusa 343)

Health - claims related to access to healthcare and new discoveries in the field (e.g., closure of emergency rooms, changes in drug reimbursement, new medications and treatments for Alzheimer's disease).

Example: **“Not because they don't want to work in emergency services, but because these services are disorganized”** pointing to the excess of patients in emergency rooms “due to a lack of support in the community and institutions.” (Text: Lusa 5)

Culture - Claims about various areas and industries of culture (e.g., state support for the arts).

Example: **“One-off support is not biennial, it does not provide the same stability that the minister talks so much about.”** (Text: Lusa 329)

Science and environment - claims related to science or scientific discoveries, both national and international, in a wide range of areas, such as genetics, archaeology or environmental issues, such as climate change (e.g., discoveries that change our understanding of history, scientific trials and research, measures to reduce greenhouse gas emissions, climate change).

Example: “**Has started operating Europe's largest photovoltaic solar energy production unit for self-consumption (UPAC) with a power output of 19.6 MVA [megavolt-amperes], comprising 43,000 panels and an energy production capacity of over 40,000 MWh per year, representing a total investment of over €20 million.**” (Text: Lusa 179)

Technology - Claims related to new technologies, both national and international, such as social media, artificial intelligence, or the internet (e.g., new legislation on user privacy, cyber espionage, digital services).

Example: “**Portugal is at the forefront of this cutting-edge technology, with advantages for everyone.**” (Text: Lusa 705)

Local - Claims related to a specific part of the national territory (changes to public transport routes, cultural heritage issues, criminal cases).

Example: “**The interregional lines that AMP would contract were defined in the 2019 agreements, so CIM do Cávado would already have been aware that this line would be terminated.**” (Text: Lusa 95)

A.2. Stance

Definition: The stance represents the position of the agent (i.e., the person responsible for making the claim) regarding the content of the *claim*. In this scheme, the stance annotation indicates whether the claimer (the person/entity associated with the claim) is affirming or refuting the content of the claim (Reddy et al., 2021).

The following examples illustrate the notion of *claim* (identified in **bold**) and stance, as understood in this project.

STANCE	EXAMPLES
AFFIRM	“ There are about two million people living on less than 591 euros per month. ” (Text: Lusa 187)
REFUTE	“ One-off support is not biennial, it does not provide the same stability that the minister talks so much about. ” (Text:Lusa 329)

A.2. Non-claim

Definition: Sentences that contain traces or indication of subjectivity, such as opinions, beliefs, information, and personal positions (Arslan et al., 2020). In addition, speculative sentences, as well as all sentences containing r future references, cannot be proven. For this reason, they are not considered factual and, consequently, do not constitute a *claim*.

Text segment: Text segments contained in declarative sentences that occur in direct speech and in quotation marks will be annotated. Quotation marks and full stops should not be included in the text segment. Sentences should be annotated individually, i.e., the annotation must be done one sentence at a time.

The following list illustrates the notion of *non-claim* (in **red bold**), as understood in this project.

Example: “**Portugal is definitely the favorite to finish in first place,**” Erich Brabec said, adding: “**In every position, Portugal has two or three good players competing in Europe. They have a lot of quality in every position on the field.**” (Text: Lusa 70)

Example: “**I believe that the decision of the President of the Republic [of Guinea-Bissau] is not correct.** It is an unconstitutional decision because it violates Article 94 of the Constitution. The article states that parliament cannot be dissolved within one year of its election,” Bacelar Gouveia (Constitutional law expert) told the Lusa news agency. (Text: Lusa 33)

Example: “**The real criminals are those who deserve to be in prison,**” he (Italian Deputy Prime Minister and Minister of Transport and Infrastructure, Matteo Salvini) added. (Text: Lusa 27)

- Claims and non-claims can exist simultaneously in the same quote.

Example: “In response, Al-Jaber stated: “**I agreed to come to this meeting to have a sober and mature conversation. I am not in any way subscribing to any alarmist discussion. There is no science, or no scenario, that says phasing out fossil fuels is what will achieve [the target assumed in the Paris Agreement of] 1.5°C.**” (Text: Lusa 251)

B. Claim Span

Definition: Within the claim (sentence), it is possible to identify the exact boundaries of the claim (Reddy et al., 2021).

Text segment: The claim span must be annotated within the claim. The *claim span* must present a complete statement with all the information necessary for a full interpretation of the sentence. Quotation marks and full stops should not be included in the text segment. Each *claim span* must be annotated individually.

In the example below, the claim (in **bold**), the claim span (in **turquoise bold**), the topic, and the stance, as understood in this project, are identified.

Example: "**Housing support is essential for teachers who, given their salaries, cannot afford housing costs**," stated Mário Nogueira". (Text: Lusa 105)

In this example, the topic is society and the stance is affirm.

Example: "**The Foupana dam, or potential dam, and the banks of the Foupana stream have no protection status, therefore they fall within the area designated as a possible prospecting area**," stated the deputy mayor, emphasizing that in the public consultation, entities could only "issue opinions based on the territorial management instruments currently in force". (Text: Lusa 34)

In this example, the topic is local and the stance is affirm.

- A sentence contains only one claim, but it may contain more than one claim span, in which case it represents more than one aspect that must be verified.

Example: "**What we know from an internal perspective is that our candidacy has managed to mobilize the overwhelming majority of our supporters, but we also know, from the most scientific methods at our disposal, that the majority of Portuguese people prefer our candidacy and our project**" he pointed out. (Text: Lusa 190)

In this example, the topic is politics and the stance is affirm.

C. Claim object

Definition: The *claim object* refers to the main point of the *claim*, taking into account its topic. Thus, the *claim object* seeks to identify what is being portrayed in the topic of the *claim*. The *claim object* is always included in the *claim span* and, whenever possible, the most specific term that provides the most context should be annotated. It can refer to both words and verbs, depending on the article. When the *claim object* is incomplete within the *claim span*, the link *Identity* must be used to link it to the expression outside the *claim span* that provides better context. Determiners are not included.

Text segment: The *claim object* must be annotated within the *claim span*. The *claim object* can correspond to one or more words, or even an expression, but it must be as succinct as possible. A *claim object* must be annotated for each *claim span*. For two *claim spans* there must be two *claim objects*. Basically, and put succinctly, the *claim object* represents what is at stake in a given *claim*. Quotation marks and full stops should not be annotated.

The following examples illustrate the notion of *claim* (identified in **bold**), *claim span* (identified in **turquoise bold**), *claim object* (identified in **pink bold**), *topic*, and *stance*, as understood in this project.

Example: “**In a context of worsening living conditions for most of the population, multinational energy companies have seen their profits increase scandalously,**” the communists note in the request.”. (Text: Lusa 38)

In this example, the topic is economy and the stance is affirm.

- There can be two *claim spans* with two *claim objects*.

The following example illustrates the concept of *claim* (identified in **bold**), *claim span* (identified in **turquoise bold**), *claim object* (identified in **pink bold**), *topic*, and *stance*, as understood in this project.

Example: “**We began Internal Medicine month with the news that, unsurprisingly, there were unfilled vacancies in the competition to choose specialties, and Internal Medicine was the hospital specialty with the most unfilled vacancies,**” laments the president of SPMI. (Text: Lusa 5)

In this example, the topic is health and the stance is affirm.

- The claim object may, sometimes, be related to elements that are located outside the textual segment of the claim span, namely through anaphoric referente (pronouns such as: *this*, *that*, *that one*).; in this case, the interpretation of the *claim object* requires referring to information from the previous context. Annotators should, therefore, identify the *claim object* within the claim, linking it to the corresponding anaphoric expression through an Identity link.

The example below illustrates the notion of *claim* (identified in **bold**), *claim span* (identified in **turquoise bold**), *claim object* (identified in **pink bold**), *topic*, and *stance*, as understood in this project.

Example: The Polish prime minister today called for the restoration of the **European licensing system for Ukrainian transport companies sending trucks to Poland**, after a month of border blockades by Polish truck drivers.

“**This system worked until spring 2022 and worked well,** then it was changed by the European Commission and today truck drivers are blocking the border,” he added. (Text: Lusa 12)

In this example, the topic is international and the stance is affirm.

- In certain cases, there may be more than one possible *claim object*. In such situations, if it is not possible to reach a consensus, the claim object is not annotated.

D. Claimer

Definition: The claimer corresponds to the entity that makes a claim. These can be, for example, public figures (congressmen, ministers, presenters, or former presidents or people in the media sphere, in general; an expert in a particular field, such as doctors or economists; or institutions, such as foundations, associations, organizations, or, alternatively, one (or more) official representative of a given institution. Identifying the claimer may be relevant in determining whether or not the statement corresponds to a *claim*.

Indeed, there is a difference between a *claim* being made, for example, by a user of a social network, such as Facebook, or by a member of the House of Representatives (or the Parliament). It is worth noting, however, that even political, business or institutional sources can be questioned regarding the veracity or factuality of the alleged content.

A final note goes to the fact that several entities associated with different *claims* may appear in the same article. Thus, whenever a claim is identified, the respective speaker must also be identified. Once the *claimer* has been identified, they should be annotated as: *person, institution, political party, organization, study, report, or investigation*.

Text segment: The text segment corresponding to the *claimer* may be, for example, the names of individuals (*Luís Montenegro*), political positions (Prime Minister), or both (the Prime Minister, Luís Montenegro) political, social, cultural, and academic institutions, political parties, organizations, unions, foundations (or their representatives). The determinant should also be annotated in cases where the claimer is associated with a company or a lawyer, for example, without any additional identifying name. In the case of studies, reports, and/or investigations, annotators should seek to ascertain whether these are reliable documents from reliable sources. If the sources of the statements raise doubts, they should be annotated as *claimers*. It is important to note that the *claimer* closest to the *claim* should be annotated, and when the *claimer* appears, for example, in relation to their profession title in another sentence that provides more context about the *claimer*, there must be an Identity link connecting the closest name to the other.

In the examples below, the *claim* (in **bold**), any *non-claim* (in **red bold**), the *claim span* (in **turquoise bold**), the *claim object* (in **pink bold**), and the *claimer* (in **green**) are identified as understood in this project.

Example: “**In a context of worsening living conditions for most of the population, multinationals in the energy sector have seen their profits increase scandalously**” said the PCP. (Text: Lusa 38)

In this example, the topic is economy and the stance is affirm. The claimer is classified as a political party.

- The same *claimer* can be referred to in different ways throughout the article, for example, by their first (or last) name or by their position. The *claimer* must be identified by the name given in the article that best identifies the entity, position, or person to whom the *claim* is associated. Note the example below. In this case, the *claim* is associated with the union specified in the first paragraph as the *Altice Group Workers' Union in Portugal (STPT)*. In this specific case, the text segment ‘the union’ should be annotated as the *claimer*, linking, through the Identity link, to the most complete claimer, i.e., ‘Altice Group Workers' Union in Portugal (STPT)’. This makes it possible to know the specific union that made the claim.

In the examples below, the claim (in **bold**), any non-claim (in **red bold**), the claim span (in **turquoise bold**), the claim object (in **pink bold**), and the claimer (in **green**) are identified as understood in this project.

Example: The Altice Group Workers' Union in Portugal (STPT) said today that the company's proposal in the collective bargaining agreement (CBA) negotiations leaves all workers “with a loss of purchasing power.”

The union recalled that, “**when it presented its proposal at the meeting on November 24, the Executive Committee stated that, with that proposal and what had been negotiated for 2023, it would be compensating Altice group workers, especially those who earn less, for the loss of purchasing power in 2022 and 2023.**” (Text: Lusa 25)

In this example, the topic is society and the stance is affirm. The claimer is classified as an organization.

Example: “We begin Internal Medicine month with the news that, unsurprisingly, **there were unfilled vacancies in the competition to choose specialties, and Internal Medicine was the hospital specialty with the most unfilled vacancies**” laments the president of SPMI.

Lelita Santos argues that, “**since Internal Medicine is the specialty that sustains hospitals and, in turn, is central to the National Health Service, it is necessary for everyone, especially policy makers, to give due consideration to this issue**”. (Text: Lusa 5)

In this example, the topic is health and the stance is affirm. The claimer is classified as a person.

- The *claimer* does not always appear before or after the *claim*; it may appear at the beginning of the article or in paragraphs located in different parts of the text

Example: The coordinator of the Bloco de Esquerda, Mariana Mortágua, said today that her party wants to “heal the wounds of the socialist absolute majority,” look at the problems, and provide solutions.

“The moment we are living in now is one of departure from a socialist absolute majority that left hospitals in chaos, housing as an impossibility for the majority of the population, and determined that many people who work have lost purchasing power,” said. (Text: Lusa 60)

In this example, the topic is politics and the stance is affirm. The claimer is classified as a person.

- If the citations are made by entities considered to be official (and, therefore, credible) sources that produce factual and substantiated statements (e.g., police or government sources), they are not considered *claims* and they are not annotated.

The following example illustrates a credible source that cannot be annotated (in burgundy), as understood in this project.

Example: In a report issued today, the ECA (European Court of Auditors) emphasizes that “competition levels in public procurement contracts for works, goods, and services have declined over the past decade in the EU single market.” (Text: Lusa 1)

- Similarly, citations referring to reports, letters, decrees, or any type of document associated with a *claimer* considered to be a credible and qualified source, should also not be considered *claims*. The following example illustrates the source (in gray), the report, which is not considered a *claimer* in this project.

Example: “The short-term outlook points to a slight acceleration in inflation in the fourth quarter of 2023,” according to the Economic Outlook and Inflation Prospects report, completed this month by the Bank of Mozambique and to which Lusa had access today. (Text: Lusa 50)

- In this project, credible sources/entities whose citations’ truthfulness cannot be questioned are not annotated. According to Gradim (2000), these sources are competent and qualified, and are therefore authorized to provide reliable and trustworthy information. On the contrary, if the source is dubious and raises suspicions as to the veracity of its statements, it is considered in the annotation.

Thus, there are several types of sources that can be taken into account in the annotation process. It is important to note that, in certain situations, the source may be anonymous, because the people or organizations involved do not want to be identified, establishing a

privacy and source preservation agreement with the journalist. In this case, they will not be considered for annotation.

1. According to Schmitz (2011), a source can be an official entity. This source refers to public figures with public functions or positions who are part of the State (e.g., the Assembly of the Republic, Ministries, Municipal Councils, Parish Councils, police and judicial forces), or aggregated organizations, such as the Institute of Registration and Notary Affairs, public companies, among others. These sources issue information of public interest and with the potential to be reported, usually through press releases. However, for Schmitz, the information can be misleading if it preserves public and political interests to benefit certain groups. As mentioned earlier, in these situations, annotators must consider the trustworthiness and credibility of sources.
2. On the other hand, a source may be unofficial. This is the case for all entities that are not related to the state, such as associations, companies, trade unions, sports clubs, among others.
3. The source may be a company (or a business), either the company itself or a representative of a particular industrial, commercial, or service company.
4. There are also institutional sources, i.e., non-profit organizations or social groups. Generally, this source seeks to raise awareness and mobilize society for social or political causes.
5. A specialist, or a specialized source, is a person or organization with in-depth knowledge in a certain area or field of expertise. This type of source is capable of analyzing events and deducing their possible consequences.
6. Ordinary people and artistic personalities can also be journalistic sources, provided they are not affiliated with organizations or social groups. They may be victims or witnesses of a particular action. This type of source is called an individual source.
7. Finally, there are reference sources, such as documents, articles, press releases, reports, newspapers, magazines, audiovisuals, websites, and portals that journalists consult and use to support their narrative and journalistic content.

E. Time

Definition: Indication of the explicit, implicit, or relative date associated with the statement of the claim. Even if this is not expressly identified, annotators should take into account the entire text to understand the time associated with the *claim*, through relative expressions such as “today,” “yesterday,” “last week,” explicit dates, such as “November 19,” or an implicit date such as “Christmas” (December 25). *Time* is only linked to a claim when it is directly connected to the moment of the *claim* (or the allegation).

Text segment: The temporal expression (explicit, implicit, and relative) associated with the *claim* must be annotated and can be found after, before or in the claim. *Time* can be found in the context surrounding the claim that the journalist writes or in the claim itself.

The following examples illustrate the notion of *claim* (identified in **bold**), the *claim span* (in **turquoise bold**), the *claim object* (in **pink bold**), the *claimer* (in **green**) and the *time* (in **yellow**), as understood in this project.

Example: **On December 29**, Nuno Dias, a resident of Almada, questioned **the mayor, Inês de Medeiros**, about "the indiscriminate felling of stone pines," indicating that it was allegedly for timber sales, and also referenced the construction of boardwalks, considering they would have an environmental impact on the reserve.

"Trees at risk of falling were cut down." **Inês de Medeiros** explained, adding that the intervention is not the municipality's responsibility, but that the Almada City Council, in the Setúbal district, "supports everything that is done to protect the Mata dos Medos." (Text: 877)

In this example, the topic is science and environment and the stance is affirm. The claimer is classified as a person.

Example: **The coordinator of the Bloco de Esquerda, Mariana Mortágua**, said **today** that her party wants to "heal the wounds of the Socialist absolute majority," look at the problems, and provide solutions.

"The absolute majority closed hospitals and emergency rooms, left Portugal in a housing crisis, and was responsible for impoverishing a vast segment of the population that lost purchasing power," **Mariana Mortágua** emphasized in statements to journalists in Almada, in the Setúbal district. (Text: Lusa 60)

In this example, the topic is politics and the stance is affirm. The claimer is classified as a person.

Example: **"Russia has once again violated the laws and customs of war, as well as the norms of international law,"** accused **Ukraine's military strategic communications center** **on Saturday**, referring to an "execution video." (Text: Lusa 299)

In this example, the topic is international and the stance is affirm. The claimer is classified as an institution.

- If there is no time expression (that is, no date is associated with the *claim*), the label *Time* isn't assigned and nothing is annotated.
- The date of publication (month/day/year) must be labeled as *Publication Time*. Whenever possible, annotators must establish a link to the respective date of publication using the *Identity* link structure.

The following examples illustrate the *time* (in **yellow**) and the *publication time* (in **orange**), as understood in this project.

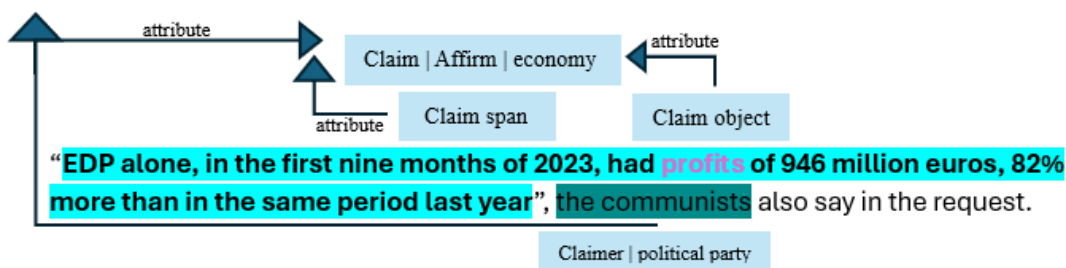
Example: Bissau, **December 4, 2023** (Lusa) – The President of Guinea-Bissau, Umaro Sissoco Embaló, decided **today** to dissolve the parliament, following the clashes on Thursday and Friday between security forces, which he considered to be an attempted coup. (Text: Lusa 85)

Link structures

A. Attribute

Elements such as *claim span*, *claim object*, *claimer*, and *time* are characteristics of the *claim*. Through linking structures, we can understand which *claim* a particular *claimer* belongs to, what subject (*claim object*) is being addressed, and the time at which it was said (*Time*). These links allow the *claim* to be linked to its respective *claimer*, which may be in previous paragraphs, for example.

Example:

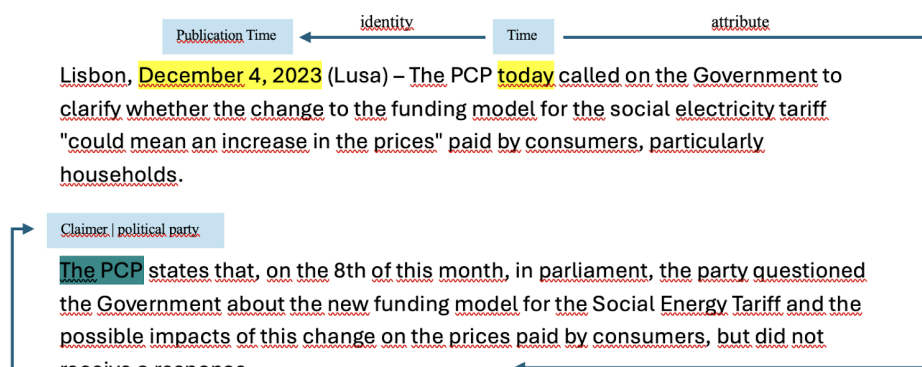


(Text: Lusa 38)

B. Identity

The *Identity* link is used to establish connections between elements that maintain anaphoric relations (through, for example, anaphoric pronouns which refer back to previous terms or expressions, directly or indirectly). This structure also allows the establishment of a link between the *time* and the date of publication of the article; a *claim object* and a *corresponding element* that is outside the *claim span*; and the *claimer*, if it is found in previous paragraphs and refers to a pronoun such as "he/she," "sir/madam," "told him/her"(that is, when the name of the *claimer* is not present).

Example:



(Text: Lusa 38)

Annotators' training phase

Before beginning the annotation process, which is carried out entirely using Inception, annotators are trained based on the annotation guidelines, which include category definitions and examples. To ensure the quality of the annotation, two trials must be conducted using a set of samples with an average of ten documents to annotate. Two annotators work on each news item.

The annotations are then presented and discussed among the annotators and with the curator in order to verify their similarities and agreements. In this way, annotators can improve their performance and possibly modify guidelines to minimize any discrepancies. This phase also provides an estimate of the average time required to annotate a text from the Lusa news agency. It should be noted that each text has its own idiosyncrasies, both in terms of length and number of *claims* and *non-claims*. After completion of the training phase, it was established that the process of annotating an article takes approximately 20 minutes. Based on this estimate and considering the weekly availability of working hours, it is expected that it will be possible to annotate approximately 105 articles per week.

However, even when guidelines for the annotation phase have been established, annotators may disagree on the classification of some sentences, assigning different categorizations to the same sentence, which may result from the fact that they have two possible classes (claim and non-claim). However, when two annotators select different options, it is not possible to define a concrete (i.e., prototypically correct) classification through voting. Therefore, in these cases, priority is given to the fact that the sentence is a claim. Therefore, in these cases, priority is always given to whether the sentence is a *claim*. At the same time, the curator can make one of the following decisions: a) if they agree with one of the classifications made by one of the annotators, they define that classification as definitive; b) the curator may also disagree with all of the classifications made by the annotators and, thus, establish their own classification. In addition to the possibility of discrepancies in the categorization of claims, annotators may not label a sentence because they assume it is non-declarative.

For these reasons, during the curation phase, the curator's role focuses on reviewing the sentences classified by the annotators. If there is a disagreement between annotators regarding classification, the curator, if in agreement with one of the classifications made by one of the annotators, can make it definitive. However, if the curator does not agree with any of the classifications, they can improve

them. And, in the event that the annotators do not label a sentence because they consider it non-declarative, and the curator disagrees, they can assume it to be a declarative sentence.