Codebook FactRank

27 March 2019

Alexander Pleijter
Peter Burger
Jan Jagers
Rafaël Hautekiet

Introduction

This codebook is intended for the development of **FactRank**: a tool that should be able to **filter checkable claims from texts**. These checkable claims are the input for fact-checkers to do their job.

Aim: establish relevant factual claims

The aim of this codebook is to reliably determine whether sentences are **checkable.** To allow this, they must be both **factual** and **relevant,** so that they are worth checking for fact-checkers.

Core guideline: assume the perspective of a fact-checker

The central guideline to follow when coding is to assume the perspective of a fact-checker: Could this sentence be the start of a fact-checking? Thus, for a sentence to be check-worthy, it should be:

1. **Factual**. That means that a sentence should contain a claim that revolves around a fact that can be checked, in other words, that it can be determined whether or not the claim is true.
2. **Relevant**. Not every factual claim is relevant for a fact-checker. Fact-checkers are only concerned with facts that matter to a broad audience. In other words, they are only concerned with claims that, if they turn out to be wrong, are reprehensibly false claims.

Coding categories

Each sentence must be assigned to one coding category. There are four categories:
1. **NF**: Not factual
2. **FR**: Factual and relevant
3. **FNR**: Factual and non-relevant
4. **Error**: Not applicable. This code is used for incomprehensible sentences. Only use this category **in very exceptional cases**!

Coding procedure

Assigning sentences to coding categories is done in two steps:

Step 1: Determine whether the sentence contains a **factual claim or** is purely **non-factual.** Factual means that it can be determined whether or not the statement is true. If the sentence is completely non-factual, choose the NF category immediately.
Step 2: If the sentence contains a factual claim, as a next step determine whether it is **relevant.** Use the following guidelines:

1. Consider: is this interesting for **a wide audience**? Do people want to know whether this is true or not? If yes, then code the sentence as relevant. Or will people think about the claim: *so what*? In that case, code it as non-relevant.
2. Pay attention to **signal words** that can make it clear that this is about a relevant fact. These could be words such as 'shocking', 'striking', 'gigantic', 'problematic', etc.
3. Pay attention to words that indicate a **change,** because changes are often relevant to fact-check. These are words such as 'increase', 'decrease', 'increasingly more/less', 'increasingly worse', 'growth', 'growing problem', etc. Of course, not every change is relevant, so try to determine from the sentence whether this is a change that is relevant to a wider audience.
4. If it is unclear what a sentence is about, then it is generally not relevant, unless you can tell from the wording that it is about something important.

What to do with unfamiliar terms?

Sometimes you come across terms in sentences that you do not know. Use Google to quickly find out what that term means, so that you can interpret the sentence better. Attention: only do this if you do not understand terms, names or abbreviations. Do not look up context or background information.

> ***Example***: The new standard for PFAS will be fifteen times stricter than the standard used by RIVM.

Explanation: A Google search shows that PFAS stands for toxic hydrofluorocarbons and that RIVM stands for the Dutch National Institute for Public Health and the Environment. This indicates that the sentence is about a danger for public health. Therefore code the sentence as FR: factual and relevant.

Guidelines for coding

In order to be able to reliably determine into which category sentences should be classified, you should use the following guidelines:

1. Multiple claims
2. Emotions
3. Opinions and value judgements
4. Predictions
5. Promises
6. Questions
7. Procedures and announcements
8. Digits and numbers
9. Reports and investigations
10. Unimportant factual claims
11. Factual claims that cannot be checked
12. Vague claims
13. Quotes
14. Proverbs and sayings
15. Performative sentences

1. Multiple claims

Some sentences are unambiguous: they consist of a single statement (example: "swallowing a woodlouse helps against hay fever"). But often this is not the case: many sentences contain subordinate clauses or similar constructions. We refer to these cases as multiple statements. In these cases, it is important to look at whether there are one or more statements in the analysed sentence that are factual and relevant.

Sometimes, there are multiple claims of which **some** are **factual** and others **not factual**. Then, because of the factual part, you have to **code the** claim **as factual**.

> ***Example***: "Both were convicted of rape at some point in time, both are outspoken, and both have had an eventful life".

Explanation: The fact that they have been convicted of rape is a factual claim that can be checked. The other two claims - being outspoken and having had an eventful life - are not factual, in the sense that you cannot determine in a clear or objective way whether they are true or false. Since the multiple assertion also contains a factual assertion, we code the whole sentence as a factual assertion. We cannot tell from this sentence who it is about, and therefore we cannot judge whether it is relevant or not. Therefore, code it as FNR: factual and not relevant.

> ***Example***: "However, we must also take into account the fact that the 2015 OECD report still places our country in the top three of all OECD countries in the ranking of the most expensive tax administrations in the world".

Explanation: In this sentence, the factual claim, namely that Belgium is in the top 3 most expensive tax administrations, stands at the end. This is a factual claim. It is also a relevant claim, because many Belgians will want to know whether their country does indeed have one of the most expensive tax administrations. Therefore, code it as FR: factual and relevant.

2. Emotions and feelings

Claims about one's **own emotions** or one's own mood are coded as **not factual**.

> *Example*: "I am very pleased with this decision".

> *Example*: "I was just so emotional".

Explanation: These are sentences that merely express how someone feels. Therefore, code them as NF: not factual.

A personal **emotion** can also be **linked to a factual claim**. In this case, code the sentence as **factual.**

> *Example*: "I am saddened by the government's plans to limit the increasing number of refugees coming to our country".

Explanation: The claim that there is an "increasing number of refugees coming to our country" is factual: you can check whether this is true. Moreover, it is a relevant claim: many people will want to know whether it is true that the number of refugees is increasing. Therefore, code it as FR: factual and relevant.

> *Example*: "I also appreciate the fact that steps are being taken to link the two databases".

Explanation: This contains a factual statement ('steps taken to link databases'), but we do not know which databases are involved. This does not seem to be something that people might find interesting, and it is therefore irrelevant. Therefore, code it as FNR: factual and not relevant.

3. Opinions and value judgements

An opinion or value judgement is someone's personal opinion and cannot be disproved. Therefore, code opinions and value judgments as NF: not factual.

Opinions are not always very clearly formulated as opinions. You have to detect this and, also in this case, code the sentence as not factual.

> *Example*: "They do not need paternalism."

> *Example*: "This is therefore a good thing."

Explanation: These sentences may not sound like opinions (they do not contain "I think/I find"), but they are expressions of opinions. You could easily add "I think/I find" to the sentence without changing the meaning: "I think that they do not need

paternalism" and "Therefore I find this a good thing." These are opinions, and you should therefore code them as NF: not factual.

Some sentences are in fact an opinion, but they also contain a factual statement. In this case, you should regard them as factual.

> **Example**: "I consider it undesirable that 80% of the migrants are unemployed."

Explanation: The sentence begins with an opinion: "I consider". What follows is a factual claim because it needs to be checked whether 80% of the migrants are indeed unemployed. This is also something that many people are likely to find interesting. Therefore, code it as FR: factual and relevant.

4. Predictions

A prediction is often an opinion, a personal vision or belief. As a rule, therefore, predictions are coded as non-factual.

> **Example**: "The government and the patient will pay for it".

Explanation: It is impossible to check whether this is the case today. That is why coding as NF: not actual.

Sometimes, however, a prediction can look so concrete that it is a fact that can be checked. For example, because calculations can be made as to whether research may have been carried out into it. Note: it is only a fact if the subject is clear and if concrete figures are given.

> **Example**: "In 10 years' time, 80% of all cancer patients will be cured".

Explanation: This is a very specific prediction which can be checked as a substantiated prognosis. You can at least check whether there are any clues to this or whether it is a totally unsubstantiated claim. That is why coding as FR: factual and relevant.

5. Promises

Promises about what someone is going to do play in the future and therefore cannot be checked. That is why we are coding them as non-factual.

> **Example**: "The next weeks and months we will work on such a programme".

> **Example**: 'Next week we are going to make that decision'.

In theory, you can check these promises by checking later whether people have kept their promise, but because it is about what people say they will do in the future, we always code them as NF: not actually.

6. Questions

In principle, a question is not an assertion, and we therefore code it as NF as a rule: not factual.

> ***Example***: "Is it true that sixteen employees of ABN AMRO distributed EUR 15 million?

Explanation: This sentence is not an assertion, but a question. Therefore, it is not a factual statement. So we code this sentence as NF: not factual.

An exception is if the question contains a clear factual assumption. In that case, a question can still be a factual assertion.

> ***Example***: 'Does the Minister agree that something must be done about this growing wave of robberies?

Explanation: This is a question, but a statement is also made, namely that there is an increasing wave of robberies. It is verifiable whether that statement is true and it is interesting for many people to know whether it is true, so we codify this sentence as FR: factual and relevant.

7. Procedures and announcements

Procedural comments and announcements are factual, but not interesting to check. They should therefore be coded as FNR: factual and irrelevant.

> ***Example***: "During the first round of questions, the coalition parties get the floor first".

> ***Example***: 'Well, next week the report will come out from the European Parliament'.

> ***Example***: 'I propose that I send it to you before the end of the year, Mr President'.

8. Digits and numbers

Phrases with figures and numbers are often factual. But not always relevant.

Sentences with figures are irrelevant when it comes to an assertion that is so obvious that it is unnecessary to check.

> ***Example***: "Everyone knows that 1 plus 1 is 2".

Some sentences with figures are factual, but not relevant to check, because they are irrelevant, do not cause controversy, do not appeal to the imagination of a large public.

> E***xample***: "Together with 122 other countries, we have requested, in the General Assembly of the United Nations, that a ceasefire be declared in Aleppo."

Explanation: You can check whether the request did indeed involve 122 countries, but for many people a few more or less will not make a difference. In addition, the fact that our country requests a ceasefire is neither controversial nor counter-intuitive.Therefore, code it as FNR: factual and not relevant.

Sometimes there is an assertion with figures, of which it is unclear what it is about and whether it is important. In that case, you have to code that it is a factually irrelevant claim (FNR).

> ***Example***: It is 2.5% in the EU and 1.8% in the Netherlands.

Explanation: Concrete figures, in other words a factual assertion, but it is not clear what it is about, nor does the sentence make it clear that it is something controversial or conspicuous. That is why coding as FNR is factual and irrelevant.

The addition of a signal word could make the above example relevant.

> ***Example***: In the EU it is 2.5% and in the Netherlands only 1.8%.

Explanation: The addition 'but' expresses that this is a peculiarity. You still don't know what it is about, but because it is put forward as a striking exception, you can assume that it concerns something relevant. So in this case coding as FR: factual and relevant.

An exception is when you can tell from the sentence that it may be something important, remarkable or grand. Or if there is a remarkable development or increase.

> ***Example***: 'Then we will have approximately 1 500 billion a year by 2050'.

Explanation: It is unclear what this is about (nothing is said about the subject), but given the amount of the amount and the fact that an increase is suggested ('then we're on...'), we can assume that it is an important subject. In this case, therefore, we codify this sentence as FR: factual and relevant.

> ***Example***: "As regards Canada, it is even 90%."

Explanation: You cannot know what this is about. However, since a number is being mentioned together with the signal word "even", it can be a relevant factual claim. [Other typical phenomena and signal words indicating FR are comparisons: "increasingly", "growing", etc.] Therefore, code it as FR: factual and relevant.

> ***Example***: "Preferably 70 to 80 cases per year".

Explanation: Figures are mentioned, but it is unclear what this is about. But the signal word 'preferably' indicates that it may be something controversial or excessive. That is why coding as FR: factual and relevant.

When figures are mentioned in a sentence that comes across as official, technocratic or specialist, we encode the sentence as irrelevant.

> ***Example***: There is an average prediction error of 3.4% for the year 2017.

Explanation: This is an example of an assertion with a figure, but with too little context to understand what it is all about. Moreover, this sounds very technocratic ('prediction error'). So we encode this sentence as FNR: factual and irrelevant.

> *Example*: The Autumn Memorandum shows that in 2018 €3.7 billion less was spent than expected.

Explanation: A concrete amount of considerable size is mentioned here, but it sounds like an official communication which is not controversial. So coding as FNR: factual and irrelevant.

9. Reports and investigations

References to reports or investigations often point to factual assertions based on something. And which can therefore be checked by consulting the reports and investigations. That is why we code such sentences as factual and relevant.

> *Example*: 'I am not saying all that; that is what health economists in various studies and the Centre of Knowledge conclude'.

Clarification: Even if you don't know exactly what this is about, you can check whether it is true that health economists conclude something in different studies. So coding as FR: factual and relevant.

10. Unimportant factual claims

If claims are factual, but you cannot infer from the context that they are about something important, then you code that sentence as irrelevant. For example, they are facts that are too technical, too official or too specialised to be relevant to the general public. Or that do not cause surprise or controversy.

> *Example*: 'We also debated this in committee'.

> *Example*: "New facts are now emerging".

Explanation: These are factual assertions, but so meaningless that they are irrelevant. So coding as FNR: factual and irrelevant.

> *Example*: "81% of Belgians have digital TV".

Explanation: This is a clear factual assertion, but digital TV is now perceived to be well established today and few people will know exactly how high this percentage is. That is why coding as FNR: factual and irrelevant.

11. Factual claims that cannot be checked

There are allegations which are factual, but which are unlikely to be properly or adequately checked by fact checkers. We are also coding them as irrelevant.

> *Example*: "However, as the person concerned has never resided in Belgium, the only link is that of a possible former spouse or partner".

Explanation: Whether someone has stayed in Belgium is in principle a fact to be checked, but it is impossible for fact checkers to determine with certainty whether someone has ever stayed in Belgium. That is why we code this sentence as FNR: factual and irrelevant.

12. Vague claims

Separate sentences that come out of a text are often rather vague: you cannot infer from that sentence what it is about. Often you can't even see what the subject is. Phrases that are too vague but contain a factual statement are coded as irrelevant.

> *Example*: 'But the population has opted for this and the Netherlands should do the same'.

Explanation: It is verifiable whether the population has 'opted in' for this, but in this sense it is not clear what that is about. That is why we codify this sentence as FNR: factual and irrelevant.

> *Example*: 'This clearly demonstrates that 2025 is feasible provided that a decision is taken now and no more time is lost'.

Explanation: Something is being demonstrated, so a factuality is being suggested. But it is totally unclear what this is about, so we codify the sentence as FNR: factual and irrelevant.

Sometimes it may seem vague, but if it can sound so relevant and also be plausible that it has been laid down in protocols or reports, then it can still be coded as FN: factual and relevant.

> *Example*: "There are far more CT scans than necessary".

Explanation: The vagueness here is in the 'more than necessary'. Where the boundary lies between 'too much' and 'not too much' is not specified in the statement. However, you can check whether there are any reports or studies that show that unnecessary CT scans are often carried out. X-rays are radioactive and unhealthy if you get too much of them, so people's health is at stake. That is why coding as FR: factual and relevant.

Sometimes an actual sounding statement is made which is not immediately clear, but which is nevertheless relevant and can possibly be operationalised in order to check whether or not it is correct.

> *Example*: "Our nuclear power plants are among the most robust in the world".

Explanation: 'Robust' as a term is rather vague, but you can sense that it refers to safety. When do experts call a nuclear power plant 'robust'? What safety criteria must then be met? What is the situation in ours compared to elsewhere? That can be checked, for example with reports from the International Atomic Energy Agency. Unsafe nuclear power stations can cost human lives, so the claim is relevant to a wide audience. That is why we code this sentence as FR: factual and relevant.

> **Example**: "The taxshift, by colleague Johan Van Overtveldt, worked on the wages of the employees".

Explanation: It is vague what is meant by 'has worked', but it suggests that there was an intended effect and that this effect was achieved. This is relevant to check. That is why we code this sentence as FR: factual and relevant.

Sometimes statements sound factual, but they are too general and vague, making them irrelevant.

> **Example**: 'Something also needs to be done about meat consumption, because that is a very big reason for pollution and that makes your ecological footprint huge'.

Explanation: This claim is too general ('very big reason') and not controversial, everybody agrees. Therefore coding as FNR: factual and irrelevant.

> **Example**: 'It is about the income of millions of people'.

Explanation: It is not clear what it is about, so it cannot be established whether it is relevant. Therefore coding as FNR: factual and irrelevant.

> **Example**: "There are now people at home with all kinds of implants in their bodies of which they do not know whether those implants have defects".

Explanation: This is a journalism relevant topic and it is a factual one, but the claim is not concrete enough for fact checking. No numbers are given. With these formulations it is enough to find two or three people with implants of which they do not know if they work. So this phrase is too vague to fact check. Therefore coding as FNR: factual and irrelevant.

> For **example**: 'All of them have received less money than promised, while there are major problems'.

Explanation: 'Received less money than promised' is a verifiable claim, but it is unclear what this is about. However, the addition 'major problems' suggests that this may be a relevant issue. That is why coding as FR: factual and relevant.

> **Example**: "These substances are found everywhere in the Netherlands, not only in Dordrecht, where Chemours is based - the company that makes these substances - but also around Schiphol Airport, Sliedrecht, Papendrecht, Helmond, Eindhoven, Moerdijk; the total list is longer".

Explanation: This sentence is not entirely clear: which substances are involved? We do not know. But Chemours (possibly searching with Google) is a chemical company, so it will be about chemical substances. That - public health! - makes it relevant to check whether those substances are indeed found in various places in the Netherlands. So this sentence is FR: factual and relevant.

13. Quotes

Sometimes someone mentions someone else's statement. It may be worth checking whether the person mentioned did indeed say that.

> **Example**: 'The Minister has said that the number of schools that fail to meet the standard has fallen'.

Explanation: This sentence contains two facts that can be checked: 1) whether the minister actually said this, and 2) whether the number of schools achieving the standard has fallen. We do not know exactly what this is about (because it is not mentioned), but this may be relevant because it is about schools and achieving standards. That is why coding as FR: factual and relevant.

> **For example**: "Delbeke said that he does not expect this to happen".

Explanation: You can check that Delbeke (possibly with Google to find out who that is) did indeed say this, but it does not appear from the sentence that it is about something relevant, so coding as FNR: factual and irrelevant.

14. Proverbs and sayings

Proverbs and sayings are not uttered in order to claim something factual. That is why we always code them as NF: not factual.

15. Performative sentences

Performative sentences are sentences that express an action that becomes reality by pronouncing the sentence. These kinds of sentences are therefore factual, but not relevant. That is why they are coded as FNR.

> **Example**: 'I call for this plan to be adopted'.

Explanation: Someone says they are making an appeal, but by saying it, the appeal is made immediately. That is why we coded this sentence as FNR: factual and irrelevant.

> **Example**: 'I advise the Minister'.