Coding Rules

COALITIONAGREE Dataset

Heike Klüver
Svenja Krauss
Hanna Bäck
Christopher Uhl

Acknowledgement:
The COALITIONAGREE Dataset was compiled with the generous support by the German Research Foundation (Grants KL 2692/1-1 and KL 2692/3-1).

Table of contents

Basic rules – Unitising and coding **2**

Unitising – Cutting sentences into pieces **2**
When to cut sentences . . . 2
When not to cut sentences . . . 3
Subcategories . . . 3

Coding **4**
The code allocation . . . 4
Ambiguity of language . . . 5
Ambiguity of Quasi-Sentences because of complexity - A hierarchy of context . . . 5
Statements containing several messages . . . 6
Statements containing no messages . . . 6
Proximity of contradicting codes . . . 6

Specific provisions - Rules to keep in mind **7**
Rare occasions: when to use the "000" category . . . 7
Catch-All categories . . . 7
Agriculture . . . 7
Background knowledge vs. personal bias . . . 8

Specific rules for the subcategories **8**
Coding: The general-category . . . 9
European and national level . . . 9

Special issues and their codes **9**
Taxes . . . 9
Transparency . . . 10
Agriculture and Farmers . . . 10
Nuclear Energy . . . 10
Women . . . 10
Immigrants . . . 11
Homosexuality/LGBT . . . 11

The following coding instructions are taken from the Manifesto Coding Instructions (5th revised edition) to foster comparability between the projects. Specific instructions have been added for the subcategories.

Two-Step Process: Unitising and Coding – Basic Rules

Basic rules – Unitising and coding

The central question of coding is: **What message is the coalition trying to convey to the citizens? Which are the issues the coalition regards as important?**
The decision-making process of coding is described in the following sections. This procedure comprises two steps: a) unitising (how many unique statements do parties make?) and b) coding (what kind of statements do parties make?).

Which Parts of a Coalition Agreement Should Be Unitised and Coded? Each textual part of a coalition agreement needs to be unitised and coded. Some parts of the coalition agreement, such as chapter or section headings, statistics, and tables of content should not be considered as text. Headlines are be coded as 800. Coders earmark those parts of the manifesto that should be ignored in the coded document. These decisions are discussed with the supervisor.

Unitising – Cutting sentences into pieces

The coding unit is a quasi-sentence. One quasi-sentence contains exactly one statement or "message". In many cases, parties make one statement per sentence, which results in one quasi-sentence equaling one full sentence. Therefore, the **basic unitising rule** is that **one sentence is, at minimum, one quasi-sentence**. In no case can two or more sentences form a quasi-sentence. There are, however, instances when one natural sentence contains more than one quasi-sentence.

When to cut sentences

Only if the natural sentence contains more than one unique argument should this sentence be split. There are two possibilities for unique arguments: 1) a sentence contains two statements that are totally unrelated; or 2) a sentence contains two statements that are related (e.g. they come from the same policy field) but address different aspects of a larger policy.

Clues to unique statements might be 1) semicolons; 2) the possibility to split up the sentence into a meaningful bullet point list; 3) general clues from codes. Regarding the third point, it is especially likely that the sentence includes two unique statements if a sentence contains codes from two or more domains. An example would be:

"We need to address our close ties with our neighbours (107.01) —— as well as the unique challenges facing small business owners in this time of economic hardship. (402.04)"

When not to cut sentences

There are many instances when sentences should not be split into quasi-sentences. A good rule of thumb is that **one word** is most likely not a quasi-sentence. It is crucial to know that **examples, reasoning, explanations**, etc. are not unique arguments and are therefore no separate quasi-sentences.

Coders should also be careful when unitising based on sentence operators such as commas, colons, hyphens, etc. Such operators might be, but are not always, indicators of a quasi-sentence. Operators do not indicate two quasi-sentences if they do not separate two unique statements. Examples for this case are:

"The animal rights in our country must be improved; and we will do that. (501.04)"
"Our country's budget must be put on solid footing again, no matter the costs. (414.02)"

Coders should not split up a sentence just because they think they have discovered a code. For instance, the mere singling out of another country is not a unique argument and, therefore, a quasi-sentence. **Only if** the statement refers to a general or specific foreign policy goal should it be considered a separate quasi-sentence. Furthermore, references to policy areas such as education, agriculture, labour, and the environment should not automatically be separated simply because 'catch words' such as 'schools', 'farmers', 'unions' or 'environmentalists' are mentioned. Again, the sentence should only be cut if it is a statement about the issue. Here is an example of a sentence that seems to contain several arguments at first glance but, on closer inspection, is revealed to have only one unique message:

"We must force our unions to step back from their demands or their policies will result in the loss of thousands of jobs, closing of schools, and diminishing pensions. (702.03)"

his example, jobs, schools, and pensions are only instances outlining the negative impact of what will happen if the coalition's central demand (unions reducing *their* demands) is not met.

Subcategories

Extending the Manifesto Project coding scheme, from which the above instructions originate, with subcategories might tempt the coder to find more quasi-sentences than one would based on the Manifesto Project coding scheme. To prevent this, it is **important not to unitise based on possible subcategories but to unitise based on the messages in the coalition agreement**.

"We need to modernize our roads and railways."

In this case, two subcategories could apply (411.02 (Roads: Positive) and 411.04 (Railways: Positive)). Such a short and general sentence should, however, not be split up into two quasi-sentences because one or two words are most likely no quasi-sentence. Thus, this statement should not be split and coded as 411.01 (General: Positive).

"Two of the main objectives of this government are to increase the coverage of our railway system (411.04) —— as well as building third lanes for our highways. (411.02)"

While the meaning of these two examples is similar, the second example is more detailed. Hence, each of these goals (roads, railways) themselves become more important, develops into a message on its own and not an example of the overall infrastructure. Thus the statement is split into two quasi-sentences and coded differently.

Coding

The code allocation

The following questions are central to the decision making of assigning codes to quasi-sentences: What are the statements of the coalition? Which policy positions does the coalition convey? In order to make this decision, coders need to make sure that they understand what the coalition says. Therefore, it is essential to read every single quasi-sentence very carefully.
Often political actors are very clear in their statements and candidly say what they seek: more of one thing, less of another. In this case, assigning codes is straightforward: coders identify the message and assign the corresponding category. When consulting the category scheme it is important to keep in mind that the categories? definitions and the subcategories are not exhaustive. They are meant to give a general notion and some exemplary statements. The scopes of the categories are not constrained to the exact wording of the definition and it should be assigned to all issues that are related to the general idea conveyed. If a statement fits into the main category but not into one of the subcategories, coders should assign the statement to the "General" subcategory. Also, if a statement fits into more than one subcategory, it should be coded into the general category.

There are, however, times when these statements are not very clear and are more difficult to code. When facing such an ambiguous sentence, the coders should always first think about the meaning of the quasi-sentence and double-check the quasi-sentence with **all** codes in the category scheme. This helps assure that the quasi-sentence does not simply fall into one of the lesser used, 'rare' categories.

In general, there are three possible factors which cause ambiguity: 1) *Language is often simply ambiguous.* Language is full of various styles, jargon, rhetorical meanings, colloquialisms, etc. Coalition agreements, therefore, often use language in manifold ways. 2) *Quite often political actors not only say what they want to achieve but also how they want to achieve it.* Sometimes, coders will find both statements within one natural sentence and will have to decide how to handle this high density of information. 3) *Many of the political issues included in coalition agreements are very complex and it is not possible to convey a clear message within one quasi-sentence.* Parties and presidential candidates often choose to build their arguments

over several sentences, within a paragraph and/or sometimes even over the course of a whole chapter.

Coders need to keep these sources of ambiguity in mind in order to fully understand the message conveyed. The following section addresses ways for coders to handle 'ambiguous' language and other problems during the course of coding.

Ambiguity of language

a) Often, political actors make policy statements by mentioning a negative aspect of an issue in order to highlight its importance. Take, for example, the following:

"Our country's democracy does not work well enough anymore!"

This sentence could be read and interpreted as a negative statement towards the country's democratic processes. However, it is rather clear that the coalition is not making a statement against democracy itself. The actual message of this sentence is one of concern about and criticism of the current state of democracy. Therefore, this is a positive statement towards the ideal principle of democracy.

b) Furthermore, political actors tend to use ambiguous or convoluted language to ?hide? certain statements often deemed politically incorrect or inadmissible viewpoints. Coders should try to understand the message while at the same time trying not to read too much into the quasi-sentence.

Ambiguity of Quasi-Sentences because of complexity - A hierarchy of context

When the quasi-sentence in itself does not convey an obvious message despite coders? best effort to find one, several levels of context might be helpful to decide how to code a quasi-sentence. These levels are hierarchical. Coders should keep in mind that it is imperative to consider the context level closest to the quasi-sentence first and only move to the next level in case the closer one was not helpful.

The context levels are, in sequence from the quasi-sentence level upwards:

1. the rest of the sentence in case the quasi-sentence is only part of a natural sentence
2. the previous and the following sentences
3. the whole paragraph
4. the whole chapter or section
5. the whole coalition agreement
6. the political discourse concerning the issue in the country at the time of the election

Statements containing several messages

Sometimes more than one code seems to apply to a quasi-sentence because the coalition wraps several statements up into one broad statement. Quite often, these statements come in the form of "We want to reach A by doing B and C" or "We are doing B and C because we want to *reach* A." In principle, the grand rule of 'code the message' applies. For these two examples, the message is that A is primarily important. B and C are simply means to achieve A. **Goals usually take precedence over means when assigning codes**. The following example claims that changing the constitution might serve the purpose of promoting animal rights. Since the constitution change is clearly only a tool, this sentence is not cut into two quasi-sentences and only the animal rights code applies.

"To make sure that animal rights are universally recognised, we are going to add them to our constitution. (501.04)"

However, there are instances where this logic does not apply. It might be possible that the coalition not only sends a message for A but also puts so much emphasis on B and C that B and C become messages in themselves. This is most apparent when the quasi-sentence states that B and C are the only means possible and there is an imperative to use them: *"We want A therefore we must employ B and C as the only feasible options."* The following example is one where the means (leaving NATO and reducing the military) are such strong messages in themselves that they need to be coded separately from the goal (peace).

"In order to achieve worldwide peace, (106) —— our country must leave NATO (105) —— and reduce the military to a minimum. (105)"

Statements containing no messages

There are instances when a sentence by itself does not make a statement. Often, the context helps in these cases and the rules mentioned above still apply. A special case is when sentences are used as a way to introduce or end an argument, or to connect two arguments. These introductory, terminal, or connecting sentences do not constitute meaningful statements themselves but are part of a continuous argument. Therefore, they should be coded in the same category as the corresponding argument or as the bulk of the paragraph in which they appear.

Proximity of contradicting codes

Finally, a note of general caution: it is possible to have positive and negative codes on the same issue right next to each other. Coalition agreements often include contradictory statements. Coders should not try to assume 'hidden' meanings in a quasi-sentence just to make sense of the sentences around it. Coalition agreements are not codes to be deciphered. Instead,

coders should be careful to only code what is written. The following examples are seemingly contradictory statements in close proximity:

"We will support our troops overseas, (104.02) —— while working to end the current war. (105.01)

Our constitution is a model for every truly democratic system (203) —— but we need to change it (204)."

Specific provisions - Rules to keep in mind

There are several rules for the process of code allocation that stem from decades of experience with manifesto coders. There are certain habits and behavioural patterns which all coders (and especially new coders) should try to avoid. Therefore, while the following rules might seem trivial, coders are asked to keep them in mind.

Rare occasions: when to use the "000" category

Generally, coders should **try to use a meaningful code (101 to 706) whenever possible.**However, there **are** instances when '000' is an applicable code. The instances are: 1) the statement is totally devoid of any meaning and cannot be coded within the context; 2) the statement refers to a policy position that is not included in the category scheme. This may be particularly true for modern issues or if the category scheme only includes codes in one direction (positive or negative) and the statement refers to the non-included direction. For instance: environment is a positive category with no negative counterpart. If a statement can only be classified as "Environment Negative" it should be given a '000' code. All quasi-sentences treated as uncodable must be rechecked after the program has been coded in its entirety.

Catch-All categories

303, 305, 408 and the 700-categories (except 703, see below) are meant to be catch-all categories for general policies that do not fit any specific coding category. They should always be avoided if a more specific policy category can be used. However, this does not mean that they are forbidden. Coders should double-check the usage of these categories to make sure they have not missed a specific policy.

Agriculture

When agricultural issues are mentioned, coders often have the choice between "703 Agriculture and Farmers" and another, often economic, category. In these instances, a special rule applies: If coders can choose between 703 or any other category, 703 should be chosen.

However, this does not mean that the inclusion of the word 'farmer' automatically makes the category 703. This category should only be assigned if the statement is actually about agriculture and farmers.

Background knowledge vs. personal bias

Generally, background knowledge is helpful and beneficial for the process of coding. However, background knowledge should not be confused with a coder's personal characteristics, beliefs, and attitudes - all of which are potentially harmful to the comparability of coding.

Background knowledge is unique knowledge that only citizens of the country can have. It includes knowledge of the country's history, social problems and cleavages, electoral issues, party system, and party ideology. Personal biases, on the other hand, are coders' individual beliefs and attitudes concerning social and political issues, party ideologies, politicians and generally concerning what is 'right' and 'wrong'.
Coders should draw on background knowledge to help determine the code of ambiguous quasi-sentences only. However, coders should only do so if no other clues are available. In all cases, personal bias must be avoided! Such bias causes distortion. Coders should be especially careful when coding coalition agreements including their most favourite and least favourite parties!

Furthermore, coders need to make sure that the statement is coded as it reads. If a coalition claims that their policy proposal has certain outcomes, this needs to be coded as it stands, even if coders think that these policies will lead to other or even opposing results. Again, the central focus of coding is to find out the policy positions and points of view of each coalition. Any personal judgements (of 'rightness' or 'wrongness', whether a statement is realistic or sensible, etc.) need to be avoided. The following sentence is a good example:

"We will increase the military expenditure to ensure peace in our region. (106.01)"

This sentence might sound incorrect but, nevertheless, the coalition is conveying the message that they want to improve the region's prospect for peace (106.01).

Specific rules for the subcategories

The previous sections covered the unitizing and coding rules from the Manifesto Project codebook. The following section will be devoted to special unitizing and coding rules for the more fine-grained codebook. However, keep in mind that we need to be compatible with the Manifesto Project coding.

Coding: The general-category

The subcategories in the codebook are not exhaustive, meaning that there might be quasi-sentences that fit into the category but not into one of the subcategories. We therefore included a general-category, where all messages can be coded into that do not fit into one of the subcategories. However, the coder should always try to find a more specific subcategory first before coding the quasi-sentence into the general category.
The general category should also be used if the quasi-sentence comprises more than one subcategory but should not be split into two quasi-sentences according to the above rules.

> *"With the assistance of the Cohesion Fund and expanded Structural Funds we will step up investment in our transport infrastructure over the next few years (411.01)."*

In this case, transport infrastructure might refer to 411.02 (Roads: Positive), 411.03 (Aviation: Positive) or 411.04 (Railways: Positive). Since it is not obvious about which way of transportation the coalition is talking, this statement is coded into the general category.

European and national level

We additionally added a variable indicating whether a statement refers solely to the national political level; or if it (also) addresses the European level. If a government talks about policies at the European level, they are not automatically coded into the two European Union categories, 108 and 110. These two categories should only be used if the coalition talks about positive or negative aspects of the EU as such (Institutions, Treaties, etc.). They should not be coded under 108/110 if they are talking about policies at the European level.

If there are statements that support or oppose specific policies at the European level, the coder should not use the European Integration codes but the code the captures the specific policy (e.g. 501 for Environmental Protection) and additionally code "1" for the variable European level. Statements about the 108/110 pertain necessarily to the European level and should always be coded as "1".

Special issues and their codes

Taxes

There is no category in the Manifesto Project dedicated to taxes. Instead, here we code whom the policy is made for. So we do not code taxes but the goals these tax statements name. This may be the goal to make the tax system more efficient (303.03) or the taxes for enterprises lower, then this is coded under 402.02 (Enterprise-friendly wage and tax policies: Positive). Tax cuts for 'the people' are coded under 503.04 (Social Equality: Positive) most of the times, but might also be coded under other subcategories if it is framed differently.

Transparency

304 (Political Corruption: Negative) is only used for issues pertaining explicitly to political corruption. If statements refer to transparency this is coded as 202.05 (Transparency: Positive).

Agriculture and Farmers

If a coalition makes a statements about agriculture, it will always be coded as 703 (703.1 for positive references and 703.2 for negative references). This category also comprises statements about farmers and agricultural commodities (e.g. wool, cattle, sugar etc.).

Nuclear Energy

There are four different subcategories for nuclear energy in the codebook. Statements pertaining to nuclear energy can be mentioned in the context of environment protection (501) or a more economic context of infrastructure and power supply (411). It is therefore important to consider where a statement on nuclear energy belongs to. If there is no explicit goal provided in the statement, the context can be especially helpful: what are the codes of the surrounding statements? What is the headline of the section?

501.05 (Nuclear Energy: Negative) and 501.06 (Nuclear Energy: Positive) should only be used if the coalition is talking about nuclear energy with regard to environmental protection. Be careful not to assume a hidden meaning but rely on how the parties actually justify their claims. An example for 501.05 could be that a coalition talks about nuclear energy threatening the environment. However, governments also argue that nuclear energy is needed because it is a 'clean' and emission-free power source. Such statements should be coded as 501.06.

If a government is talking about nuclear energy as a power source as such, it is coded either under 411.12 (Nuclear Energy: Positive) or 411.13 (Nuclear Energy: Negative). Statements about nuclear energy being a cheap and reliable power source, are coded as 411.12. Negative statements about nuclear energy as being a cost inefficient and non-reliable power source are coded under 411.13.

Women

Statements about women can be coded into different (sub-) categories. If the government is talking about measures **against the discrimination** of women or **equal treatment** of men and women, this should be coded under **503.02** (Gender Equality: Positive). Very general and favorable mentions of women can be coded as **706.02** (Women: Positive) but only if this statement does not include specific policy positions (e.g. 503, 504, 505, 402, etc.).

Women could also be mentioned in the context of a traditional (603.02) or progressive (604.02) image of the family and in the context of the question whether to legalize abortion and birth control and/or to extend these rights (604.04). If governments are against the legalization of abortion and birth control, this is coded under 603.04 (Abortion and Birth Control: Positive).

Immigrants

Statements revolving around questions of immigration and integration can be classified into two broad categories. **601.02** (Immigration: Negative) and **602.02** (Immigration: Positive) are concerned with the possibility of **new immigrants** in the coalition's country. Hence, if a coalition agreement includes positive (negative) statements about new immigrants entering the country, this should be coded as 602.02 (601.02).

607.02 (Immigrants: Diversity) and **608.02** (Immigrants: Assimilation) include statements about immigrants that are **already in the country**. Statements, that favor multiculturalism and diversity in a country, should thus be coded as **607.02** (Immigration: Diversity). Statements that require immigrants to give up their cultural traits and language and completely adapt to their new home country's customs, is coded as **608.02** (Immigration: Assimilation). If a statement is about the equal treatment of immigrants and locals, it should be coded as 503.05 (Racial Equality: Positive). Only if statements on immigration do not fit into the above categories can such very general and favorable mentions of immigrants be coded as 705.05 (Immigrants: Positive).

Homosexuality/LGBT

If a coalition argues that the discrimination of homosexuals as such is bad and that homosexuals should have equal rights, it should be coded as **503.06** (Equal rights for Homosexuals: Positive). However, if a coalition argues that homosexuals should not be allowed to marry because a 'real' marriage, as designated by god or nature, can only exist between a man and a women, it should be coded as **603.04** (Homosexuality: Negative). If a coalition agreement contains statements against the assumption that a "real" marriage can only be between a man and a woman and thus want to open marriage for homosexuals, this should be coded as **604.04** (Homosexuality: Positive). Only if statements on homosexuals/LGBT persons do not fit into the above categories can such very general and favorable statements be coded as **705.04** (Homosexuals: Positive).