Glossary

Adapting coding from grounded theory: a data-driven strategy for generating categories, adapting open coding from grounded theory. It involves three steps: (1) identifying concepts in your material, based on similarities and differences; (2) grouping similar concepts into categories; (3) generating structure by distinguishing between main categories and subcategories.

Blind coding: see coding.

CAQDAS: computer-aided qualitative data analysis software.

Category: Categories are the building blocks of coding frames, each category corresponding to a relevant meaning. Categories can be main categories or subcategories.
> Main category (dimension): captures aspects of meaning on which your analysis is focused.
> Residual category (miscellaneous category): functions as a container for unanticipated information that is relevant to your research question, but is not described by any of the other, substantive (sub)categories in your coding frame.
> Subcategory: captures what is said with respect to your main categories.

Co-occurrence (of categories): Categories are said to co-occur if the same unit of analysis has been assigned to these (main) categories or if adjacent units of coding have been assigned to these categories. Examining your findings for co-occurrences is part of data exploration towards presenting your results in qualitative style.

Coder: person who does the coding in QCA.

Coding: The term has a broad meaning in qualitative research in general and a narrow meaning specific to QCA. In the broader sense 'coding' is used as an umbrella term to refer to a variety of methods for analysing qualitative data and establishing links between data, other data, and concepts. In the context of QCA, 'coding' refers to that step in QCA where you assign a unit of your material (a unit of coding) to one of the (sub)categories in your coding frame.
> Blind coding: coding by two coders where neither is aware of the codes assigned by the other.
> Conceptual coding: A type of coding in the broader sense that is used to 'open up' your data. It is analytical and is aimed at establishing links both between data and concepts and between concepts.
> Double-coding: coding the same units twice. Typically this is done by two coders who work independently from each other, i.e. neither is aware of the codes assigned by the other (blind coding). It can also be done by having one person code the same units twice in succession, with one or two weeks in between the two rounds of coding.
> Main coding: applying your coding frame to all your material.
> Reductive coding: A type of coding in the broader sense that reduces large amounts of material to a few general themes. It is descriptive and is aimed at establishing links between data. It is also called 'indexing'.
> Trial coding: trying out your coding frame on part of your material, proceeding exactly as you will during the main coding; this corresponds to the first step of the pilot phase.

Coding frame: the main tool in QCA for describing relevant meaning and a way of structuring your material. It consists of main categories and subcategories for each main category specifying relevant meanings concerning this aspect.

Coefficient of agreement: coefficient that expresses the degree of consistency between two rounds of coding. They can refer to either inter- or intra-rater agreement. Common coefficients

of agreement include: percentage of agreement, Scott's pi, Cohen's kappa, and Krippendorff's alpha.

Comparative coding sheet: a matrix where you enter to which subcategories a given unit of coding was assigned during the two rounds of coding, where rows represent units of coding and columns represents rounds of coding (i.e. coders or points in time). This is used in the pilot phase and during the main analysis phase. During main analysis, the sheet should contain a third column where the final meaning of each unit is entered.

Complexity: The complexity of a coding frame increases with the number of dimensions and the number of hierarchical levels. Coding frames can be simple (i.e. low complexity), of medium, or of high complexity.
> High: Coding frames of high complexity consist of several dimensions and more than two hierarchical levels.
> Medium: Coding frames of medium complexity either consist of one dimension and more than two hierarchical levels; or else they consist of several simple coding frames 'strung together'.
> Simple: Simple coding frames consist of one dimension and two hierarchical levels.

Concept-driven (deductive): working in a concept-driven way is to base one's work on previous knowledge. In QCA, using a concept-driven strategy is to base your categories on previous knowledge. This can come from various sources: a theory, prior research, everyday experience, logic, an interview guide.

Consistency: criterion for assessing the reliability of coding frames by comparing two sets of coding of the same materials, either across persons or points in time. The more consistent the two rounds of coding are, the higher the reliability of the coding frame.

Construct validity: see validity.

Content validity: see validity.

Contrasting: a data-driven strategy for generating subcategories, especially suitable for comparing material from different sources. It involves the following steps: (1) Identifying similarities within the first source and creating subcategories; (2) identifying similarities within the second source and creating subcategories; (3) identifying differences between the two sources and modifying the subcategories so that they capture these differences.

Criterion validity: see validity.

Data-driven (inductive): working in a data-driven way is to let your categories emerge from your material. Data-driven strategies for generating categories include: progressively summarising your material; adapting coding from grounded theory; subsumption; contrasting.

Data matrix: a matrix where the rows represent your units of analysis and the columns represent the categories of your coding frame. Creating the data matrix at the end of your main phase of analysis involves transforming your results from the level of units of coding to the level of units of analysis.

Decision rules: the fourth part of a category definition. Decision rules tell the coders which of two overlapping categories to use. They should specify what is not to be included in a category, which category to use instead; negative examples can also be included. Decision rules are optional and should be specified only if categories overlap.

Deductive: see concept-driven

Defining: the fourth step in building your coding frame. Defining your categories is to specify the rules for assigning data segments to the categories in your coding frame. Category definitions have four parts: a name, a description of what you mean by that name, examples, and decision rules (if needed).

Description (category description): The second part of a category definition; description of the features of a category and of any indicators, i.e. aspects of the data pointing to that category.

Dimension: see category / main category

Discourse analysis: a set of qualitative methods for analysing the ways in which language and its use contribute to the construction of social reality.

Double-coding: see coding

Examples (of a category): the third part of a category definition; units of coding which are typical of the meaning described by the respective category. Ideally they are taken from the material you used for building your coding frame; if this does not contain any good examples, examples can also be hypothetical.
> Negative examples: can be part of decision rules. Negative examples illustrate what is not to be included in a category.
> Positive examples: the default type of example. They illustrate what falls within a given category.

Exhaustiveness: requirement that each unit of coding can be assigned to at least one subcategory in your coding frame, i.e. that all relevant meanings in your material are in fact covered by your coding frame.

Expanding: the sixth and final step in building a coding frame. When expanding your coding frame, you include additional material and check whether your frame adequately describes this. If it does not, you generate additional (sub)categories.

Face validity: see validity.

Generating: the third step in building a coding frame; this refers to creating subcategories for main categories.

Grounded Theory: a qualitative approach aimed at generating theory that is grounded in the data. Key characteristics include an iterative, cyclic, and inductive procedure where case selection, data collection, and analysis are closely linked. Data analysis is carried out by means of a three-step coding procedure. Its first part can be adapted in generating subcategories in QCA.

Hierarchical level: subdivision of categories into subcategories into sub-subcategories, etc. A coding frame contains at least two hierarchical levels: main categories and subcategories, and it can in principle contain any number of levels.

Indexing: see coding / reductive coding

Inductive: see data-driven

Inter-rater reliability: see reliability

Intra-rater reliability: see reliability

Latent content: content that is less standardised, indirectly expressed, and conveys more than one meaning at a time; its meaning is usually not immediately obvious, and different persons may disagree on the exact meaning.

Main analysis phase: a phase of QCA that includes: preparing for the main coding; doing the main coding; comparing codes; deciding on the final meaning of all units of coding; and transforming results from the level of the unit of coding to that of the unit of analysis.

Main category: see category

Main coding: see coding

Manifest content: content that is comparatively standardised, direct, and conveys only one meaning at a time; its meaning is usually obvious, and different persons are likely to agree on this meaning.

Mutual exclusiveness: requirement to conceptualise and define the subcategories within any one dimension so that any unit of coding can be assigned to one of these subcategories only.

Name (category name): the first part of a category definition; a label providing a concise description of what the category is about.

Pilot phase: The pilot phase involves trying out and modifying your coding frame before using it on all your material. It consists of the following steps: (1) the trial coding; (2) a consistency check; (3) adjusting your coding frame.

Progressively summarising: a data-driven strategy for generating categories. It involves four steps: (1) paraphrasing all relevant passages; (2) deleting all superfluous information from the paraphrases; (3) integrating similar paraphrases. The third step is repeated until you reach the desired level of abstraction.

Qualitative content analysis (QCA): a data analysis method for systematically describing the meaning of qualitative material through classifying relevant passages as instances of the categories of a coding frame.

Quantitative content analysis: a method for systematically describing the meaning of verbal or visual material through classifying relevant parts as instances of the categories of a coding frame. The method overlaps with qualitative content analysis. However, quantitative content analysis focuses more on manifest meaning, follows a more linear sequence of steps, sets greater store by reliability as a quality criterion, and is considered a method of data collection, not of data analysis.

Scope of inference: distance between the textual meaning and the reference point of an inference made on the basis of analysing the textual meaning. In QCA, inferences have been made to the communication situation, the communicator, and the recipient, requiring increasingly wider 'leaps' of inference and raising increasingly grave concerns about validity.

Reliability: a criterion that is used to evaluate the quality of research instruments. An instrument is considered reliable to the extent that it yields error-free data. In QCA, reliability is assessed through consistency, by comparing coding across persons or across points in time. Consistency can be quantified by calculating a coefficient of agreement.

> Inter-rater reliability: assessing reliability by comparing how two or more persons code the same units independently from each other.
> Intra-rater reliability: assessing reliability by comparing how the same person codes the same units at two different points in time.

Residual category: see category

Revising: the fifth step in building a coding frame: Here you look at main categories and subcategories from a structural point of view, checking for overlaps between categories, mixing of dimensions, structural completeness, etc.

Saturation: requirement that each subcategory in a coding frame is used at least once during the analysis, i.e. that no subcategory remains 'empty'.

Segmentation: Segmenting your material is to divide it into units such that each segment/unit fits into one category of the coding frame.

Segmentation criteria: criteria according to which you decide where one segment ends and another begins.
> Formal segmentation criteria: make use of the structure inherent in your material. Examples include: chapters, sentences, clauses, paragraphs, words.
> Thematic segmentation criteria: make use of topic changes. Each unit corresponds to mentioning/discussing one theme.

Selecting: the first step in building a coding frame. It refers both to deciding where to start and deciding which parts of your material are in fact relevant to your research question.

Semiotics: the analysis of signs and the ways in which cultural meaning is generated through using signs.

Statistics: provides instruments for the description and analysis of empirical results.
> Descriptive statistics: allow you to describe your findings in selected respects; absolute frequencies and percentages are examples of descriptive statistics used to present the results of QCA in quantitative style.
> Inferential statistics: allow you to determine whether any differences between groups are so large that they are unlikely to have occurred by chance.

Structural completeness: refers to a set of categories that are mutually exclusive. The set is structurally complete if all mutually exclusive (sub)categories are included in the set.

Structuring: the second step in building a coding frame. This refers to deciding upon the main categories / dimensions around which you will build your frame. To do so, you can use any of these three strategies: a concept-driven, a data-driven, or a mixed strategy, combining a concept- and a data-driven approach.

Subcategory: see category

Subsumption: a data-driven strategy for generating subcategories for a given main category. It involves the following steps: (1) examining your material until you come across a relevant passage; (2) checking whether this is similar to or different from other passages already examined; (3) if it is similar, you mentally subsume it under an already existing subcategory; (4) if it is different, you create a new subcategory. These steps are repeated until it is no longer necessary to add new categories.

Text matrix: a matrix whose cells contain text instead of numbers. They are used to present results in qualitative style, allowing you to summarise and at the same time illustrate various aspects of your findings.

Trial coding: see coding

Typology: the result of sorting your data into groups so that the cases making up one group are similar to each other, and the cases that go into different groups are different from each other. Creating a typology is a useful way of summarising the results of QCA and presenting your findings in qualitative style.

Unidimensionality: requirement that each dimension in your coding frame should capture only one aspect of your material.

Units: parts/segments of your material
>
> Context units: that portion of the surrounding material that you need to understand the meaning of a given unit of coding.
>
> Units of analysis: the units selected for QCA; the units yielding one text each; often correspond to 'cases'.
>
> Units of coding: those parts of the units of analysis that can be interpreted in a meaningful way with respect to your categories and that fit within one subcategory of your coding frame. You have to mark units of coding in your material.

Validity: a criterion that is used to evaluate the quality of research instruments. An instrument is considered valid to the extent that it captures what it sets out to capture. A coding frame is valid to the extent that the categories adequately represent the concepts under study.
>
> Construct validity: refers to the relationship between the concept under study and other concepts; it is assessed by deriving and testing hypotheses concerning these relationships.
>
> Content validity: extent to which an instrument covers all dimensions of a concept; useful in assessing the validity of concept-driven coding frames.
>
> Criterion validity: relationship between an instrument and other indicators of the concept in question whose validity has already been established.
>
> Face validity: extent to which an instrument seems to measure what it is supposed to measure; useful in assessing the validity of data-driven coding frames.

Variability: an important criterion in selecting the material for building your coding frame and for doing a trial coding. Variability refers to differences in your material, usually due to different sources. Material should always be selected so as to include as many differences as possible.