Dear All,

Wojtek and I had now two long sessions on our coding categories. We have several observations and suggestions (for your convenience we include also a Word version of our remarks).

I. Recoding and reclassifying

1. There are too many codes, so we need to prune and consolidate:
   1.1. There are codes whose usefulness is not clear, therefore we suggest that an abductive method needs to kick in now and this round of pruning and streamlining of our codebook should be primarily driven by a deductive form of reasoning from our hypotheses and general question. We have employed a lot of induction so far and we seem to have a good "feel" for our data, but it is time, we suggest, to prune and discipline the set of codes.
   1.2. Relying on theories and/or conceptual models, starting with hypotheses, and adding "deduction" to the repertoire should not be mindlessly rigid, of course. My preferred method, I have learned mostly from reading the great British "interpretive" historian, archeologist and philosopher of science, Collingwood, is to focus on questions. What are our questions? So, not deductive or inductive reasoning per se, but rather a relentless search for the clarity in asking questions.
   1.3. So, for example, what is the main area of life in which people spot problems? This takes us directly to the Inglehart/Norris problematique: culture or economy. Or, more broadly: culture, economy, politics, or society (as in the Manifesto).

2. It seems that we are particularly interested in answering the following question: ***How do people see the world and – in particular – which political forces they are inclined to support because of this particular way of "seeing" the world?***
   2.1. As we were asked to "clean" the category "problems" it dawned on us that the theory of social movements and protest politics directs our attention to "problems" as the driving force behind mobilization. The existence of problems is not sufficient to generate mobilization/protest, but in some theories it is a necessary condition. In any case, it seems useful to assume together with the demand-side theories of populism's rise, that "problems' are one of the central categories of our work, which is based on an assumption that (many) people diagnose the world to be out of whack and search for solutions (also political).
   2.2. We should focus – Wojt and I suggest – predominantly on three dimensions of the people's "picture of the world" (that is their "culture" or "collective intelligence"):
      2.2.1. Problems: state of affairs or a situation that exist but should not
      2.2.2. Needs: (desired) states of affairs that do not exist but should
      2.2.3. Affirmations: states of affairs that do exist and should exist
      2.2.4. Each of them should be a Z category (we have already two of them!), but our systematic inspection of the existing codes reveals that some of them are miscategorized. Try our method: when you are looking at a fragment of the text you need to code, ask yourself whether you see there:
         2.2.4.1. Problem
         2.2.4.2. Need

2.2.4.3. Affirmation
2.2.4.4. None of the above

3. We further believe that the proper coding of "problems" requires making several key distinctions and decisions:
   3.1. Personal vs general (social, economic, cultural, political). Eg: Jirka's "pessimism." Is it that a particular person feels pessimistic or, rather, it is a person diagnosing the existence of pessimism in the society?
   3.2. Let's focus on several basic distinctions. For example, cultural vs economic will help us contribute to the question posed by Inglehart and Norris: is it culture or economy that drives support for populism and populists?
   3.3. Given the significance of "problems" (if we all agree), we will need to carefully recode some annotations and/or reclassify some codes. For example, some codes in "social and political processes" ("aging") may be presented as problems in the coded text. The first annotation on Amelia's list of 14 annotations coded as "aging" reads: "Population aging, alcoholism, suicides and other social problems are common in marginal areas of the Alps." Putting aside the problem (signaled below) that it is not "our" code, it is a "**probem**"! Explicitly named so.
   3.4. In general, **some** codes may be need to be reclassified as belonging to "problems." That will require some additional work, but we believe at that it would be worthwhile, as it would help us to focus our whole analysis more sharply around "problems" and "needs."
      3.4.1. Another, important example. Take the code "car." Why do we need to know if people talk about "car" in our project? It seems only if someone sees "car" as a problem (too many cars, so we have bad pollution) or "need" (we need more cars so people can get to work more efficiently). We could get to the idea of "car" as a "problem" buy creating an edge between two separate codes: "car" and "problem." ***The problem is that we do not have a code called "problem"!*** And perhaps for a good reason: it is awfully abstract. Solution: code "problems with…" directly. All these codes will end up in the Z category are "problems."
      3.4.2. BTW, after a careful review, I have a proposal of reducing 273 or so "problem codes" to 40. Alternatively, these 40 new codes may form another layer in the hierarchical structure.

4. Several months ago, we decided to start reconstructing what we called the **grammar of action**. Accordingly, we created such Z categories as:
   4.1. What people do:
      4.1.1. Actions and activities (1+1791)
   4.2. Who are the doers?
      4.2.1. Institutions/organizations (We are in the process of dividing this category into institutions and organizations, as we decided) (1+1030)
      4.2.2. Movements (need to separate from "historical events") (1+157)
      4.2.3. People and identity (1+408)
   4.3. How do they do this?
      4.3.1. Institutions (to be created in separation from organizations)
   4.4. What drives them?

       4.4.1. Problems **(1+1707) Observe how much is already happening here! The highest number of annotations.**
       4.4.2. Needs (for now "resource needs) (1+804)
       4.4.3. Ideology (2+677)
       4.4.4. Values and beliefs (1+1221)
       4.4.5. Emotions (3+977)
   4.5. Context of action:
       4.5.1. Social and political processes (1+922)

5. We understand that we can look for co-occurrences at five different levels (this is answered in the threat already)
   5.1. The whole Poprebel corpus
   5.2. A single thread
   5.3. Specific motifs (sub-threads)
   5.4. A single post
   5.5. A single annotation

6. Another issue we would like to discuss is this: We can isolate the Poperbel corpus by using the "discourse" function, but when we inspect a specific code, for example, "disability" we see every instance when this code was used in any Edgeryders' project, present of past. This is confusing as these posts can be totally irrelevant. Why does it matter? We believe we need to review and revise many codes. This task will be much easier and faster if the codes we will be inspecting belong only to Poprebel.

7. Should we code questions? Most likely not. Need to discuss.

II. To split or not to split (the posts), that is the question!

1. Assume three sets: posts (A, B, C… Z), codes (x, y, z), and edges (1, 2, 3, …n) and consider coding post A. There two choices. It can be coded as a whole or as a set of sub-posts (A1, A2, A3, etc.).
2. Posts are full interviews with one person.
3. Sub-posts are created by breaking down posts into segments. Each segment is composed of one question and answer.
4. Full post coding:
   4.1. Code x (Church) and code y (government) are mentioned in post A. An edge appears between the two and is assumes the value of 1.
   4.2. Assume, this code appears in 23 more posts in the whole corpus we code. If so, the edge (x, y) assumes the value of 23.
   4.3. Observe that in this case *it does not matter* if the interviewee A made an explicit connection between x and y. It is enough that the word "Church" and the word "government" appeared *anywhere* in the text of A.
   4.4. This is the problem (?) of *indirect edges*. Postulating an edge between two codes (concepts) that have not being intentionally connected in the A's (and others') argument/statement/answer can be defended only on the grounds of an assumption that relying on an unrepresentative sample of interviews we are mapping a mental map of the

world that members of a certain community (Poles, Chechs, Germans) carry in their heads.

    4.4.1.   In this approach to the "cartography", we are assuming that the two concepts (codes) are connected because they were **mentioned** in a text called "interview transcript," not because they were **explicitly connected** by the interviewee. We seem to be assuming that "mentioning" of two concepts (represented by their codes) in a single interview indicates that there is some form of a link between them.

    4.4.2.   In some if not most interviews, there will be subsets of edges resulting from the explicit connection between x and y by A, as in the sentence: "The Church should get out of government's business". Let's call these ***direct edges***.

5. Sub-post coding:

    5.1. This type of coding is designed to avoid the problem (if it is a problem) of ***indirect edges.***

    5.2. But here is another problem. If x (Church) and y (government) appear in post A, but never together in one answer (sub-pots), the whole of sub-posts of A (A1, A2, etc.) will generate no edges between x and y. We lose any information about indirect edges, but – perhaps – this is what we want.

    5.3. The result will be that our graphing device will receive information about far fewer edges than under the full post coding. Again, perhaps this is what we want, but we need to thoroughly discuss this issue.