🌻 Coding and recoding – Dealing with overlapping labels

So assuming you're not coded with a fixed, coding with a fixed codebook, and assuming you're putting in enough text that it's either coding multiple sources at a time or it's, sorry it's coding multiple sources or it's you've got one or more texts which are so long that you're breaking them up into multiple chunks, then it's normal that you'll have a codebook with very many labels, many of which overlap in meaning and so you've got the problem that we have discussed a fair amount that you can address either with soft recoding or magnetic relabelling, which we've done a lot, but we're also moving more nowadays to hard recoding where once we've done a fair amount of coding, or all the coding, we then group those labels either pre-clustering them using embeddings or and or applying an AI and or putting a human in the loop to get a list of labels for hard coding, or rather a system because you might have a smaller set of labels and additional tags or columns.

We haven't done the research to find out whether having say 10 labels and three tags is more or less efficient than the corresponding set of 60 labels.

And of course there's also the newer possibility of using the AI in the AI Answers feature to take an existing large set of coded labels and then recode them into a more compact set. This has also been discussed here

So you've basically got a lot of decisions to take as you work your way through the coding pathway. And it all depends on lots of different things like how long your text is, how many different documents you have. Oh, I didn't mention that in Causal Map we never, we do break down long source texts into smaller chunks but we never combine smaller source texts into larger chunks, so each source is always coded on its own. So if you don't have a fixed codebook then you'll always get many labels with overlapping meanings.

	Hard coding	**Hard recoding**	**Links recoding**	**Factors recoding**	**Soft recoding**
Accuracy	Highest	...	...	...	Lowest
Speed	Slowest	...	...	...	Fastest
Manual	Just code manually	Make a copy of your file, delete links and start again	Edit manually in Links table or Map, - or use	Edit manually in Factors table or Map, - or use	-

			search/replace in Links table	search/replace in Factors table - or Bulk Edit	
AI	Just code with AI, with/without a codebook	As above, or just put the switch "skip coded sources" to off	AI Answers / Links. Writes into whichever Label Set is active in the toolbar.	AI Answers / Factors. Writes into whichever Label Set is active in the toolbar.	Apply magnetic labels in Soft Recode filter

What's the point of Links and Factors recoding? What's the difference?

- Soft recoding is only as good as the underlying embedding space, and it is never perfect.
- Hard recoding can take a long time, is expensive, and does not encourage experimentation
- With Links/Factors recoding, you can:
  - Recode just the currently filtered sources/links (or all links)
  - Recode into whichever Label Set is active. Pick `default` in the Label Set widget (toolbar, below the Sources bar) to write to the permanent cause/effect labels; create a named set such as `experiment1` and the AI writes to `cause_experiment1` / `effect_experiment1` instead. Flip between sets in the same widget to view either.
  - There is also another option Answers which is not about recoding; it is simply a way to send your links and/or factors data to an AI and getting a text answer.

But the main point is that rather than just hoping the magnetisation will work the way you want it to, you can do smart recoding as if you had an assistant to work through each label. For example you can say "Relabel everything which expresses a decrease or lack of something with a ~" or "Look at all these labels and tag each with `[Food]` or `[Health]"

.

- You can even bring other columns into play, for example citation count, source count etc.
- Links recoding is significantly more powerful because you can also include the actual Quote as well as both Cause and Effect. This means the AI can make its decision with a lot more context. So this is almost like recoding from scratch, but the original coding has already identified causal claims and all we have to do is relabel the labels with the same complete information about the claim.
- Be careful: it's tempting to say things like "Find 3-8 top-level factor labels which cover the meaning of all these labels and recode them with the new top-level labels", but remember the "Rows per call" slider: with a large set of links and lots of quotes you will probably have to break up your work into multiple chunks, and each call may come up with different labels. In this case you could use the Answers mode (or the Cluster part of Soft Recode filter) first to develop some labels.

Soft recode: a useful contrast

Soft

See also

- Bulk relabelling factors for the manual routes (search/replace, Bulk Edit) in detail.
- Translate factor labels for a worked AI Answers / Factors example using a `french` Label Set.
- Recoding labels temporarily for safe experimentation with the Label Set widget.