The structure of the thesis could be based on a distinction between a methodological part and an analytic part. Roughly, the initial methodological part would consist of three chapters (this is straightforward):

– the distributional workflow (similar to Chapter 3 of the monograph)
– the visualisation tool (similar to Chapter 4 of the monograph)
– the materials used (corpus data and annotation, selection of lexical items).

The analytic part could also consist of three chapters, where each chapter would be concentrating on an important synthetic observation or conclusion that arises from looking at the results. (This is a very important point: emphasising the *essence* of your findings through the organisation of the chapters.)
As far as I'm concerned these main points would be as follows (I'm deliberately describing them in a propositional form, to highlight their status as essential arguments).

1. **The clusters that come out of an analysis do not take the form of neatly delineated sets that one might naïvely expect, but rather take the form of different nephological shapes. These shapes result from the interplay of different parameters: how a model selects different context words, how these context words related to the target into each other et cetera.**
   Basically, this corresponds to the 'technical interpretation' of your existing classification and analysis of clusters as clouds.

2. **The clusters that come out of an analysis do not straightforwardly correspond to senses but may have different types of informational status. Predominantly, these clusters may represent: idiomatic combinations, substances of larger senses, senses as may be found in dictionaries…**
   This classification of what comes out linguistically is something that you would need to extract from what is now your 'semantic interpretation' of clusters as clouds, but importantly, the discussion should not be organized on the basis of the cloud types but should focus on the classification of the different types of information that you get.
   Practically speaking, mapping annotated sentences on the cloud types would obviously support this type of analysis.
   At the same time though, cloud types might play a role, when you consider the question: what is the probability of finding certain types of information in certain cloud types? This is a way of linking up this second major conclusion with the first one.

3. **There is no optimal solution across items; what works well for one item does not necessarily do so for others, and the same holds for semantic phenomena like the ones that we used to select the set of items under consideration.**
   No further comments needed at this point; this we have discussed before.

Finally, the thesis might conclude with a practical part, or at least a practically oriented concluding chapter, listing practical guidelines for how to use the workflow and visualization tool, specifically also emphasizing how this could be useful beyond a classical corpus linguistic analysis based on a list of collocations. (As we mentioned before, what does this ad beyond the SketchEngine level?)