Peacekeeping Operations Corpus:
A use case on missions' multidimensionality

Elio Amicarelli

PKOC0A05

Contents

Summary

Peacekeeping multidimensonality, that is the progressive expansion of peacekeeping domains of activity, is often cited by the literature but has never been examined from a quantitative standpoint. In this report we show how this topic can be investigated using the Peacekeeping Operations Corpus. More specifically, we use the corpus for several tasks:

    a) Calculating indexes of missions' complexity as a proxy for multidimensionality.

    b) Investigating how peacekeeping complexity has changed over time.

    c) Investigating the evolution of activities within specific missions.

Data preparation

Our initial data are sections titles extracted from PKOC associated with the metadata of the respective report. We decided to focus on titles because they provide a concise description of the main topics discussed in a section.

Our starting set of titles is cleaned by *a)* removing entries not concerning relevant mission's activities, *b)* further cleaning the titles by keeping only the words describing an acitivity domain, and *c)* stemming the remaning words. The following example illustrates this process on a small sample of titles:

a) introduction; status of deployment of the mission; economic regeneration; reconstruction and development; demobilization and reintegration

b) ~~introduction; status of deployment of the mission~~; economic regeneration; reconstruction and development; demobilization and reintegration

c) econom reconstruct demobilizat reintegrat

As can be seen, each title is reduced to its most informative part about the domain disscussed in the corresponding report's section. The plot below shows the 10 most common domains extracted from reports' titles.
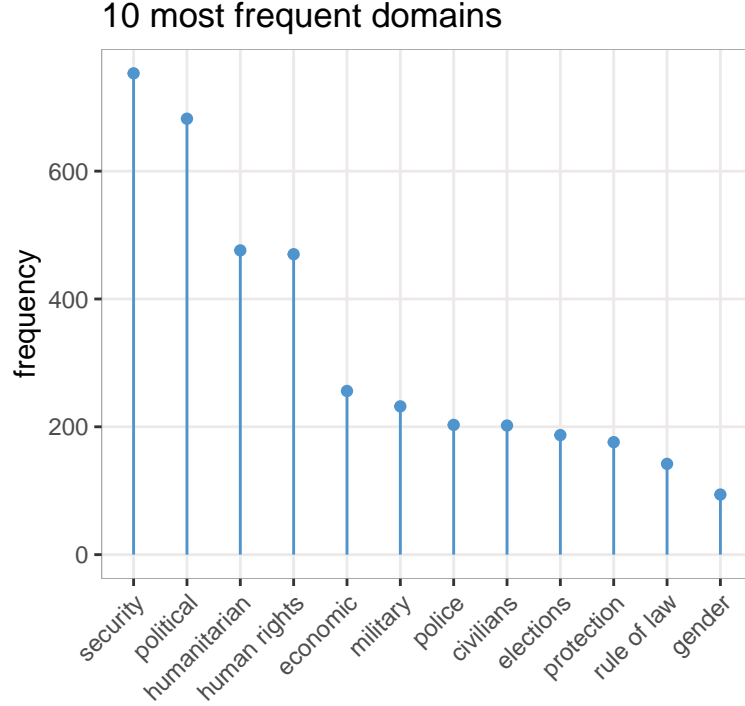

Figure 1: 10 most frequent domains identified from reports titles

Analysis

Operational definition of missions' complexity

After the data preparation stage, the content of our documents look like the following example from the BNUB mission:

peac democr sec stabl transit just govern institut law human_rights soc econom gend child protect polit sec econom peac democr govern human_rights transit just jud institut gend child protect sec peac polit sec leg elect vot reg candid environ med observ peac democr sec stabl transit just govern institut law human_rights soc econom gend child protect peac democr sec stabl transit just govern institut law human_rights integr soc econom gend child protect peac democr sec stabl transit just govern institut law human_rights soc econom gend child protect

Because of this parsimonious structure, on the one hand grammar and words ordering no longer represent informative elements. On the other hand, words occurrences and multiplicity are now the main data aspects of our interest. For these reasons, we consider Bag of Words (BoW) models as a suitable representation of our documents. We leverage the BoW model to calculate Shannon's information entropy as measure for document complexity. Shannon's information entropy for a document X is defined as:

$$H(X) = -\sum_{i=1}^{n} P(x_i) log_2 P(x_i)$$

Where $P(x_i)$ is the probability of a domain (e.g. 'econom') to appear in document X. The figure below shows Shannon's information entropy for all missions started after 1989.
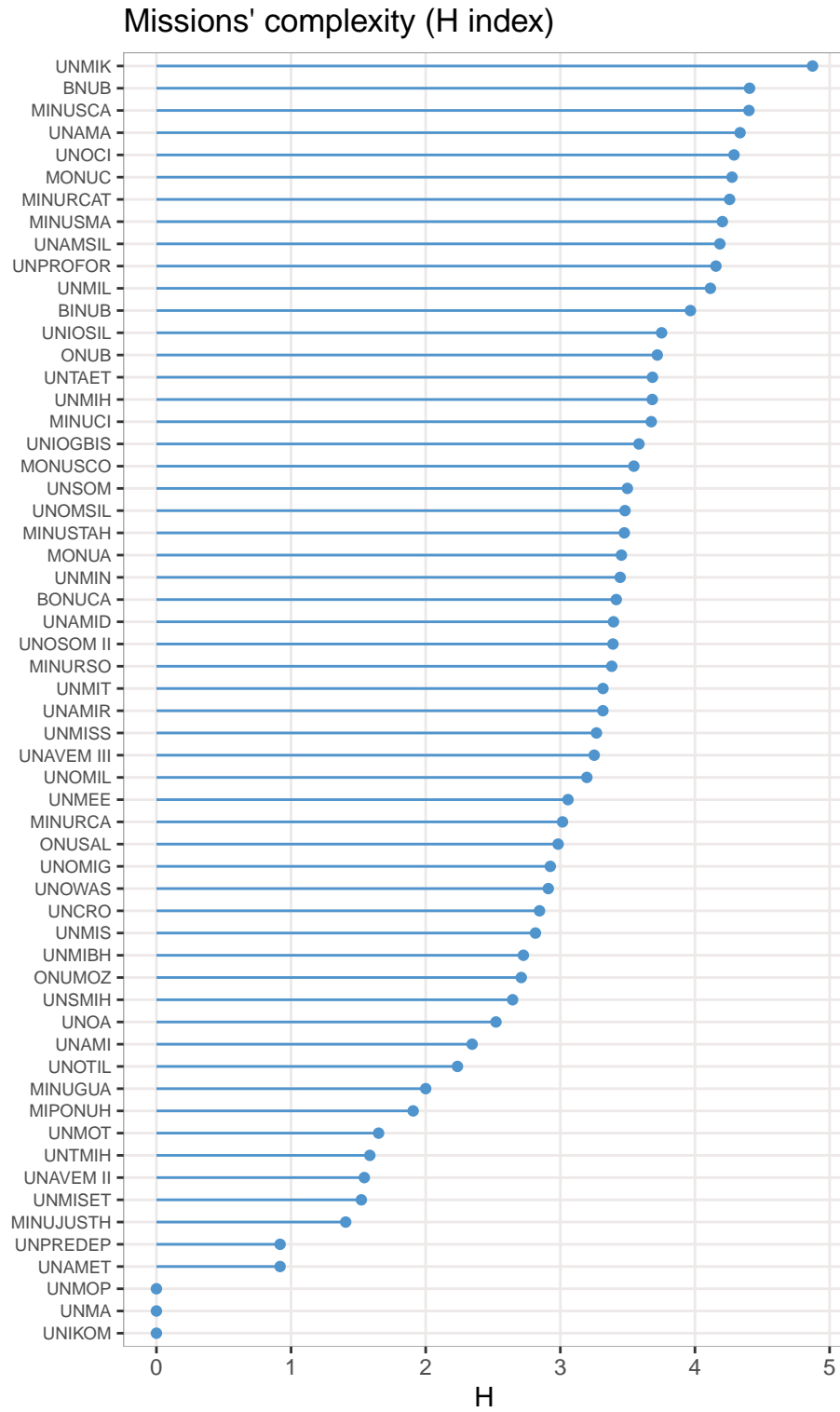

Figure 2: Information entropy on missions

The picture below uses two missions as example in order to better show how Shannon information index (H from now on) relates to documents' complexity.
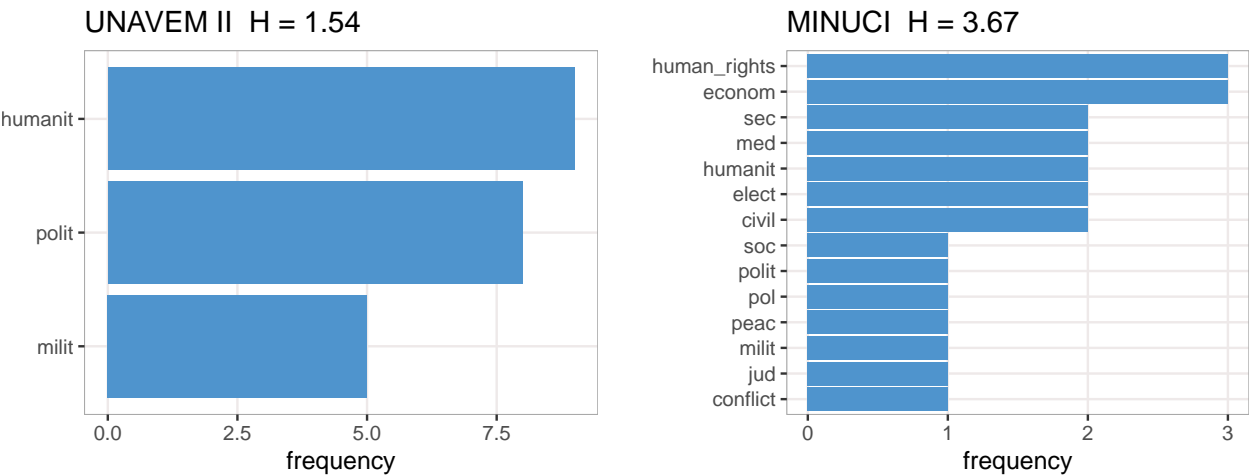

Figure 3: Domains in UNAVEM II and MINUCI. MINUCI is more complex than UNAVEM II.

The complexity of peacekeeping over time

Peacekeeping literature often mentions that peacekeeping missions have experienced significant expansion of their activities over time. We used the data and the definition of complexity described in the previous sections in order to examine this argument. By calculating the H index over the cumulative set of titles reported over the years we found support for this statement. As showed in the plot below, the H index has consistently increased since 1994 meaning that the cumulative set of activities domain reported has expanded over time and became less predictible from the persepective of information theory.
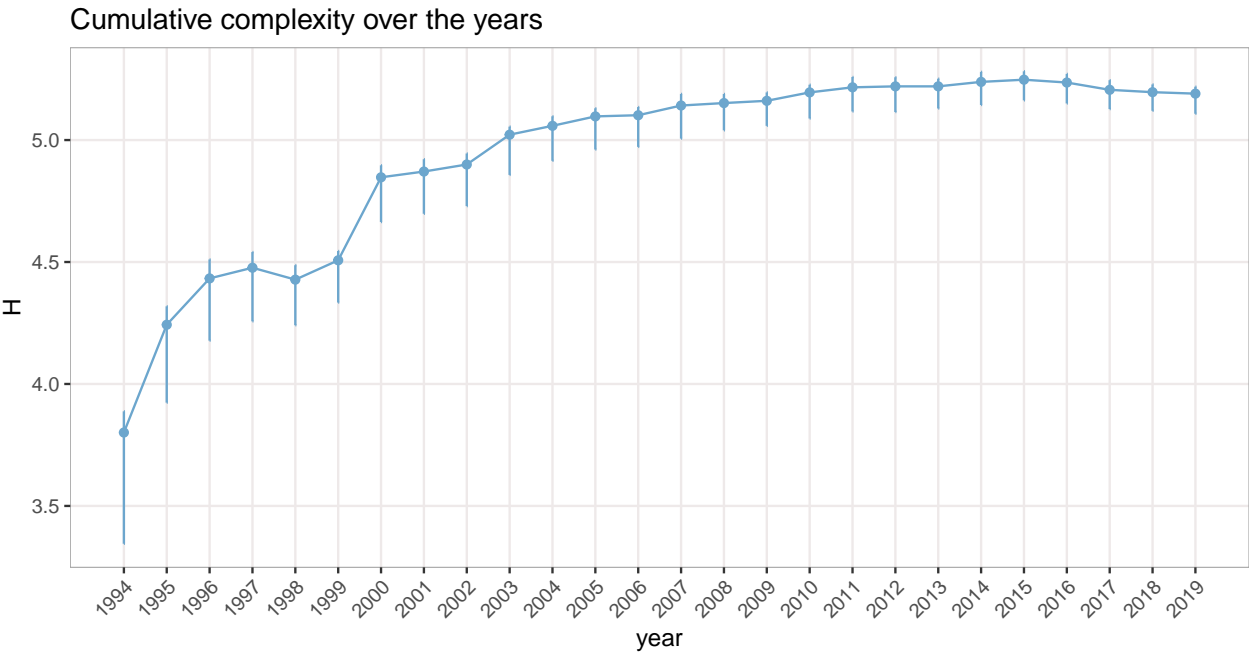

Figure 4: Complexity of peacekeeping over time

The data can also be used to track which domains of activity are more present in a given point in time. Each series in Figure 5 shows the yearly reporting proportion for a given activity domain relative to all the activity domains reported in the same year. When we look at temporal evolution for some of the most relevant domains, we see for example that activities in the security and political domains have become more prevalent during the last 15 years while activities labeled as economic or military have experienced a relative decrease.
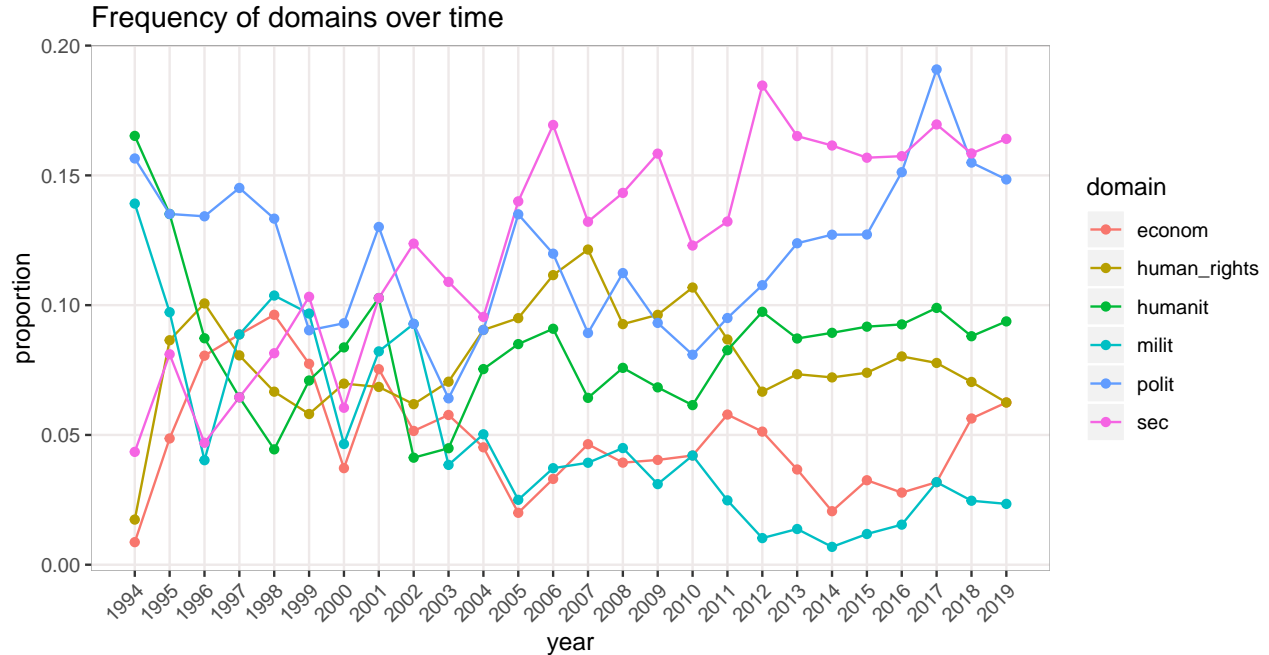

Figure 5: Activities over time

The evolution of UNMIK activities

We now move to investigating the structure of activity domains for an individual mission. As showed in the previous section UNMIK has the highest H score among all the missions we considered. For this reason we use UNMIK as an example. The plot in the figure below shows the UNMIK activity domains reported in each period where reports are available (top panel) and the evolution of the entropy index H (bottom panel).

When we look at the top panel we can see that in its initial phase UNMIK performed activities in a very wide range of domains (red box). However, in its mature stage according with the reporting the mission narrowed the scope and focused on two main branches (blue boxes) where the religious, cultural, rule of law, security and human rights became the most frequent reported domains.

As can be seen from the bottom panel, the H index provide a good summary of the evolution in the complexity of the mission: it increases rapidly in the first phase where the mission started to operate in a wide range of domains and stays more or less stable since then.
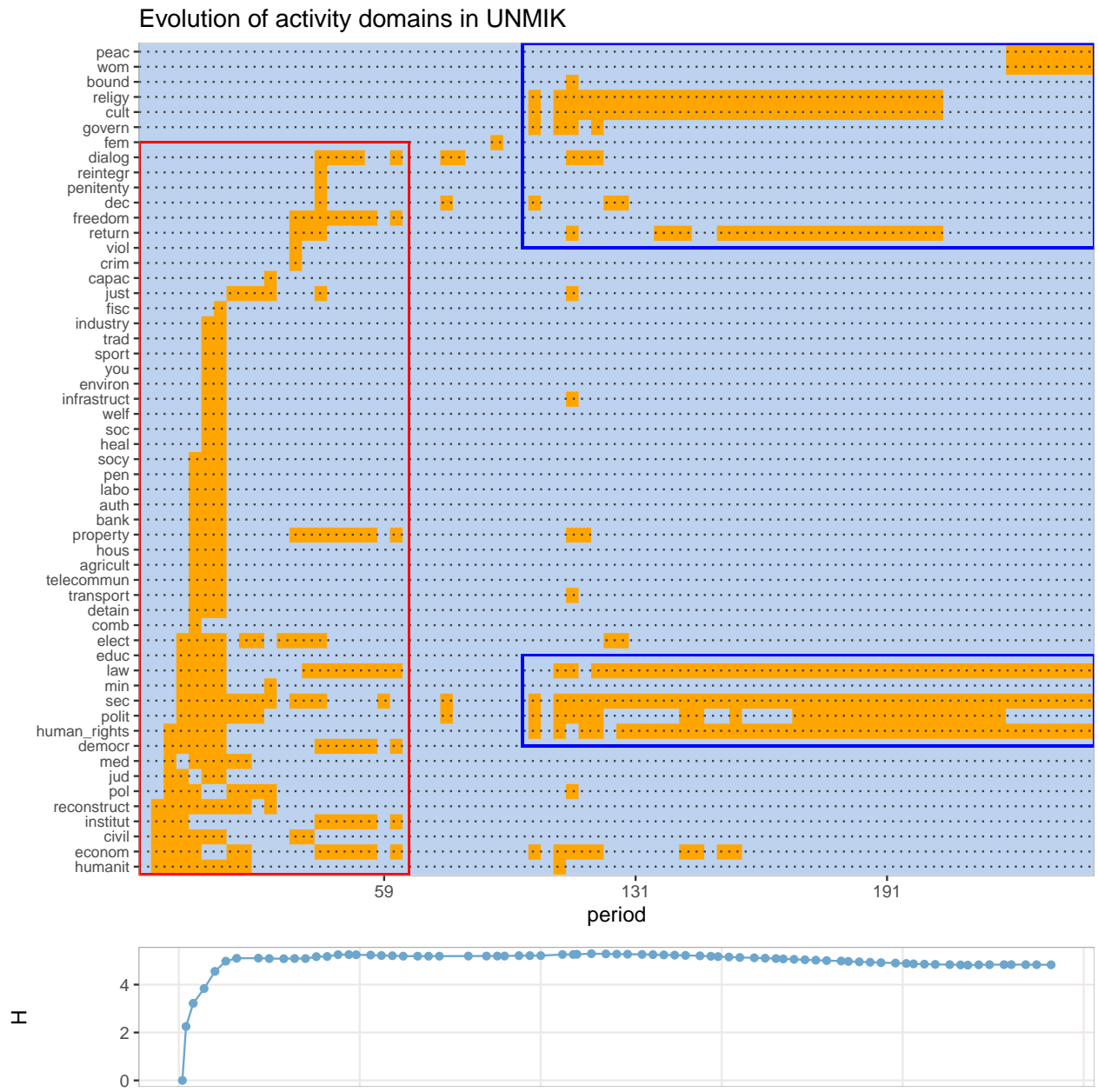
Figure 6: The evolution of UNMIK

Concluding comments

The Peacekeeping Operations Corpus offers great opportunities for researchers and policymakers. It allows to investigate aspects of United Nations peacekeeping that it wouldn't be possible to investigate otherwise given the current data landscape in peacekeeping research. In this report we showed how the corpus can be used to extract information regarding missions' complexity, the evolution of missions activities, and to obtain insights on multidimensional peacekeeping. This type of applications, together with the automated extraction of event data are just a few examples of the great potential of this new data-product.