[101] *The sources and analysis of language data in the Linguistic Survey of India*

Aimée Lahaussois (Histoire des théories linguistiques, CNRS).

The Linguistic Survey of India (henceforth LSI) is the publication resulting from a vast linguistic enterprise led by G.A. Grierson: begun in 1894, the final volume was published in 1928, the entire work covering 723 linguistic varieties from South Asia.

For many of the languages in the LSI, the entry is made up of the following elements: -an introduction to the people -census data -authorities and provenance of data -a grammatical sketch -language specimen(s) Comparative word lists, grouping all languages within a given section, are also provided.

In this presentation, I focus on a number of topics surrounding the data used in the preparation of the LSI.

a) the choice of text specimens and word lists: For almost all languages with a text specimen, the Parable of the Prodigal Son is featured. Is this because of Grierson's personal background, as suggested by (Majeed 2019), or was the choice related to the use of the same specimen for other large-scale projects? I also look into Grierson's selection of Campbell's word and phrase list (Campbell 1874) rather than other lists available at the time for the same linguistic area.

b) the process of data collection and annotation: Grierson describes some details relating to the data collection process, but I explore what can be deduced about the collection, segmentation and annotation of data by comparing the different language samples. I also attempt to understand the influence on data collection and annotation practices of an earlier work by Grierson (Grierson 1897) circulated as an example of how to prepare the data.

c) the representativity of the text specimens as sources for a grammatical sketch: How representative are the specimens in conveying important features of the grammar of the languages? The grammatical sketches that result from the analysis of the specimens and word lists probably do not stand up to Grierson's ambition of providing materials that will allow colonial administrators to learn the local languages. I examine whether this is because of the materials they are based on, or the challenge of producing sketches for hundreds of languages without having carried out fieldwork.

Campbell, G. 1874. Specimens of Languages of India including the Aboriginal Tribes of Bengal, the Central Provinces, and the Eastern Frontier. Calcutta: Bengal Secretariat Press. Grierson, G.A. 1897. Specimen Translations in Various Indian Languages. Calcutta: Government Printing, India. Majeed, Javed. 2019. Colonialism and knowledge in Grierson's Linguistic Survey of India. London: Routledge.