|  |  |  |
| --- | --- | --- |
| **System** | Task | MT, ST, fully automatic subtitling etc. |
| System Input/Output and granularity | video, audio, text; subtitles, sentences, documents, timestamps; pre-/post-processing |
| **Evaluation criteria** | Type | Quality, conformity, accuracy, productivity, etc. |
| Aspect | Content (translation), form (length, segmentation), time (reading speed, spotting) |
| **Data** | Language(s) | Language (variety) |
| Size | Clips, minutes, sentences, tokens, etc. |
| Format | Video, audio, plain text, xml, srt, etc. |
| Content | Genre, scripted/unscripted, single/multi-speaker |
| **Participants** | Number | Per language pair/total |
| Type | Expert, how related to the authors, paid/offered course credit/voluntary, etc. |
| Expertise | Language skills, subtitling skills, PE skills, expertise with the tool/interface |
| How recruited | Agency, online, etc. |
| Remuneration |  |
| **Experimental setting** | Type | PE-based, rating, error annotation etc. |
| Tool/user interface | Description and functionalities, video display, audio available |
| Requirements | Timestamps considered/ignored, format aspects and constraints |
| Scale or other rating instrument | Size, list or range of possible response values, how presented to evaluators |
| Information given to evaluators | Training, warm-up sessions, guidelines, instructions, verbatim question/prompt/etc. |
| Experimental conditions | in the lab, remotely, on crowdsourcing platform, etc. |