Appendix 1

1. Data from University of Tsukuba Hospital

The patients were randomly selected from the 53,246 patients who visited one or more of the 11 departments at University of Tsukuba Hospital between January 1, 2013 and September 30, 2018. The progress notes used here were the same as the existing dummy progress notes dataset. The data were limited to the patients who visited 11 departments because their basic characteristics were suitable for the existing dummy progress notes dataset. The actual progress notes used for this study were written in the Subjective, Objective, Assessment, and Plan (SOAP) format [14]. Notes that did not follow the SOAP format were excluded from this work, because we needed to decompose the progress notes semantically.

2. Existing dummy progress notes dataset

We used existing public dummy progress notes [16] for comparisons and evaluations. We used the notes of only 29 dummy patients, as these followed the SOAP format and could be decomposed semantically for this study. Among these notes, the assessment (A) and plan (P) were concurrently described as the assessment and plan (A/P) in the progress notes of 22 dummy patients. Therefore, A and P were concurrently treated as A/P in this study for these progress notes and for the progress notes from University of Tsukuba Hospital. In all, 257 dummy progress notes of 29 dummy patients were decomposed into the subjective data (S), objective data (O), or A/P categories, and 514 decomposed dummy progress notes were generated.

Information about the medical departments of medical the doctors who wrote the existing public dummy progress notes was not disclosed in the dataset. One doctor (the first author) evaluated the department for each dummy progress note. There were 11 departments included, namely cardiology, respiratory, gastroenterology, nephrology, hematology, collagen diseases, stroke, endocrinology, digestive surgery, oncology, and breast thyroid endocrinology surgery.