

Thanh Thieu, PhD¹, Jonathan Camacho Maldonado, MD², Pei-Shu Ho, PhD²
¹Oklahoma State University, Stillwater, OK, USA ²National Institutes of Health, Bethesda, MD, USA

Problem Statement

Identify mobility related functioning concepts in clinical notes amidst the incompleteness of the ICF¹ as a vocabulary source

Abstract

Goals

- Analyze entity hierarchy of mobility concepts in physical therapy (PT) notes
- Manually annotate a gold standard corpus of mobility mentions and validate inter-annotator agreement (IAA)
- Build machine learning models to automatically identify mobility mentions

Method

- Inter-disciplinary analysis of 1,554 PT notes
- Manual annotation using GATE Developer
- Named entity recognition with an ensemble of CRF and BiLSTM-CRF

Results

- A hierarchy of 5 entity types, 3 relations, 8 attributes, 33 categorical values
- A gold standard corpus (GSC) of 14,281 nested entity mentions on 400 clinical notes
- GSC has high IAA with 92.3% F1-score on exact matching of entity mention textual spans
- Our ensemble model achieves state-of-the-art performance of 83.31% F1-score on exact entity matching

Functioning

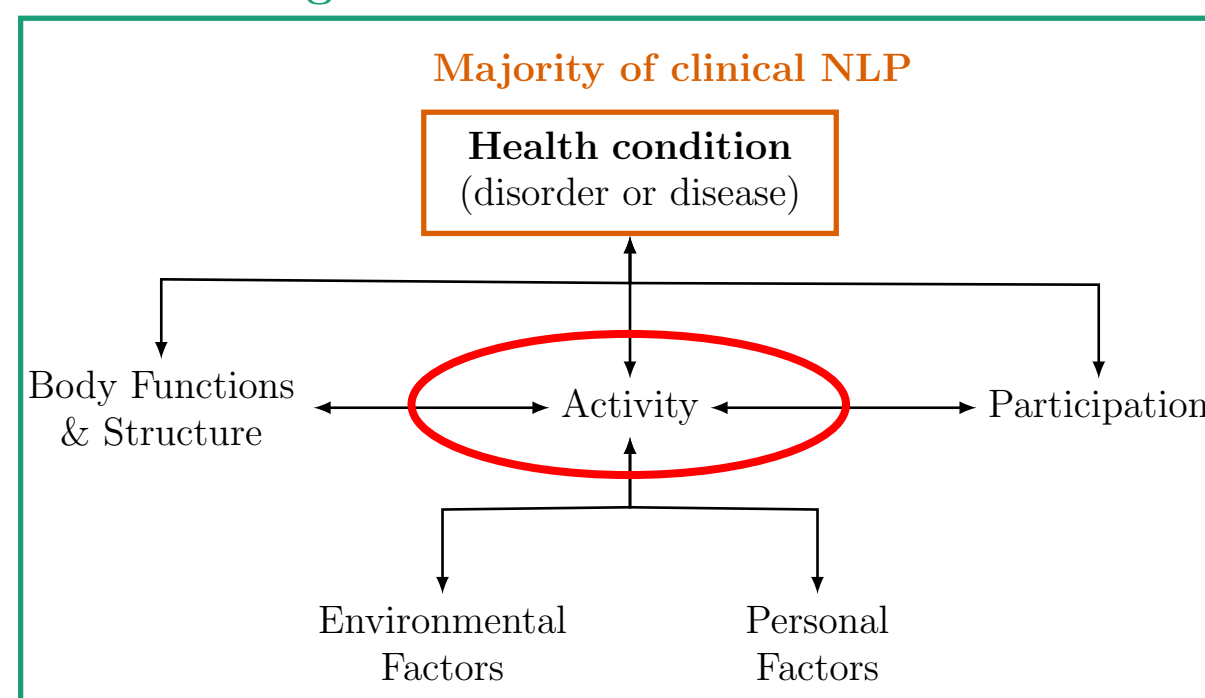


Figure 1: ICF¹ diagram of human functioning, in which Activity component is underexplored. Reproduced by permission of World Health Organization (WHO), from ICF¹, p18

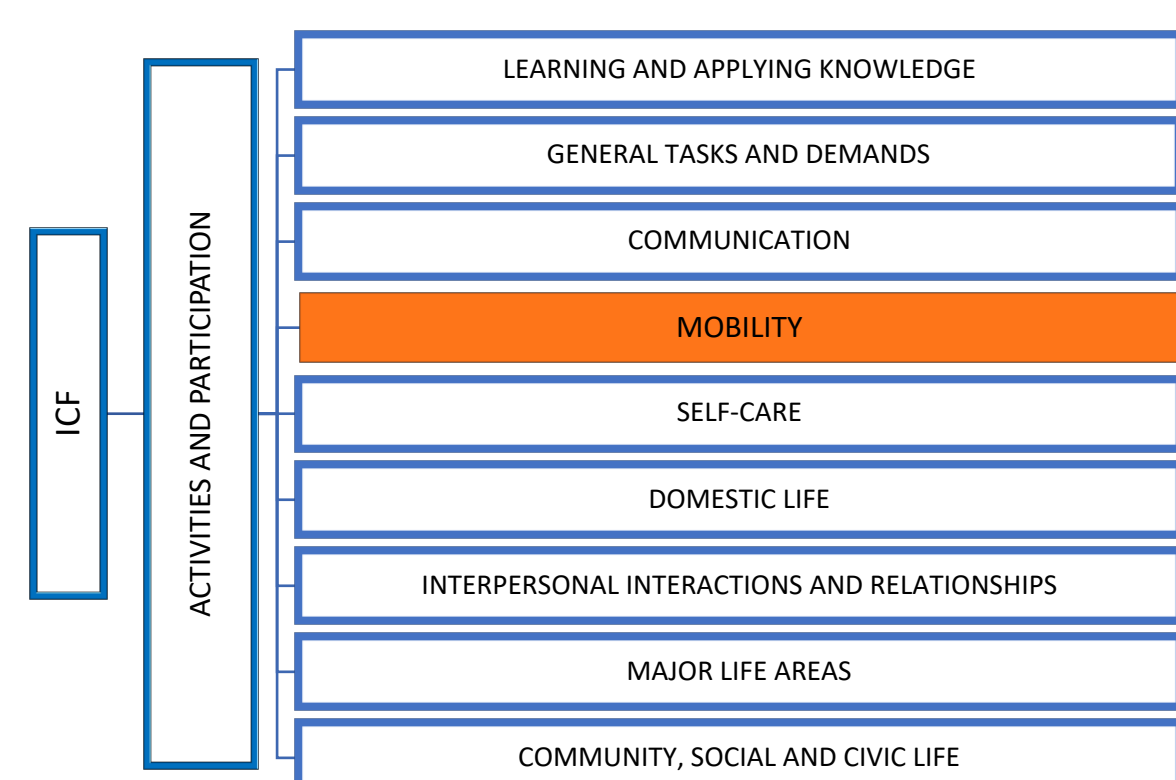


Figure 2: Mobility domain of Activities and Participation component of the ICF was analyzed for its semantic structure

Data Analysis & Annotation

- NIH Biomedical Translational Research Information System (BTRIS)²
- 577,735 de-identified clinical records of 19,005 patients from the NIH Clinical Center, total 305 note types
- 1554 PT notes contain most mobility information, including 950 PT initial assessment, 320 PT reassessment, 278 PT assessment and discharge, and 6 PT discharge
- 100 PT notes were seeded for interdisciplinary analysis of entity hierarchy, annotation schema, and annotation guidelines
- 1,454 PT remaining notes were annotated and 400 were selected for gold standard annotation

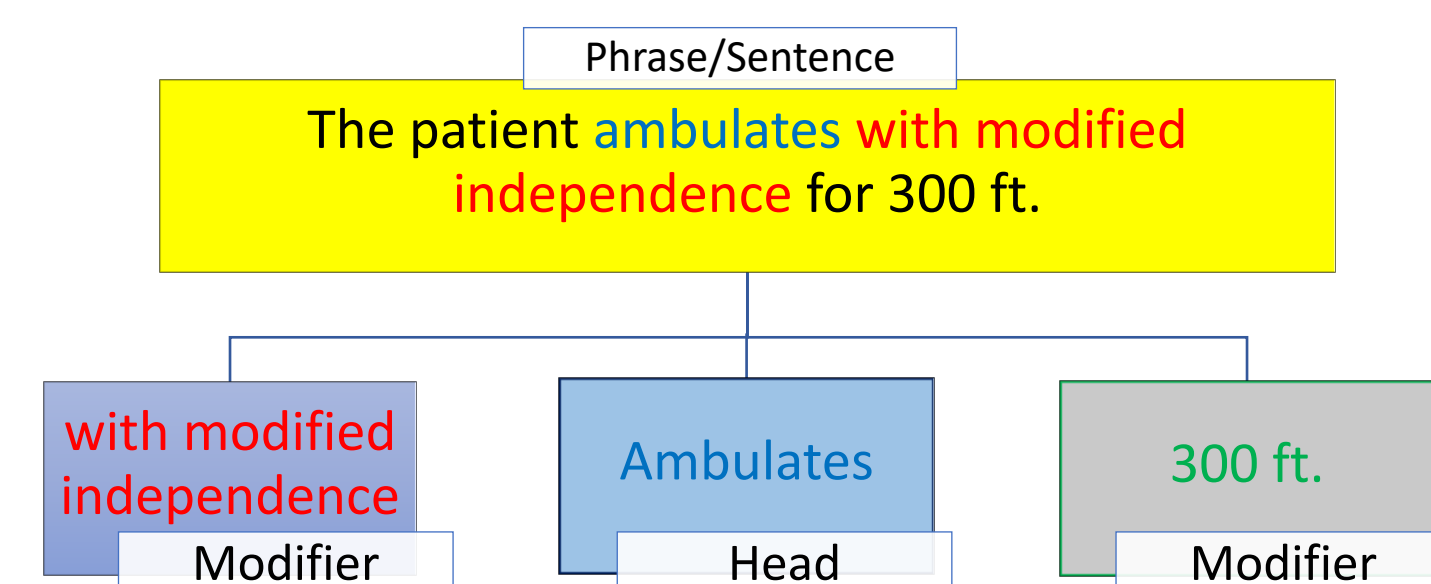


Figure 3: Semantic components of mobility concept in clinical free text

Entity Hierarchy

- Mobility: a self-contained, well-defined description of physical functional status information
 - Action: information about an activity related to mobility represented by a 3-digit code from the Mobility domain of the ICF
 - Assistance: information about dependence on another person or device when performing the activity
 - Quantification: information regarding measurement values of the activity
- Score Definition: a standardized assessment of functional status, often represented as numerical values that provide semantic interpretation of functional status

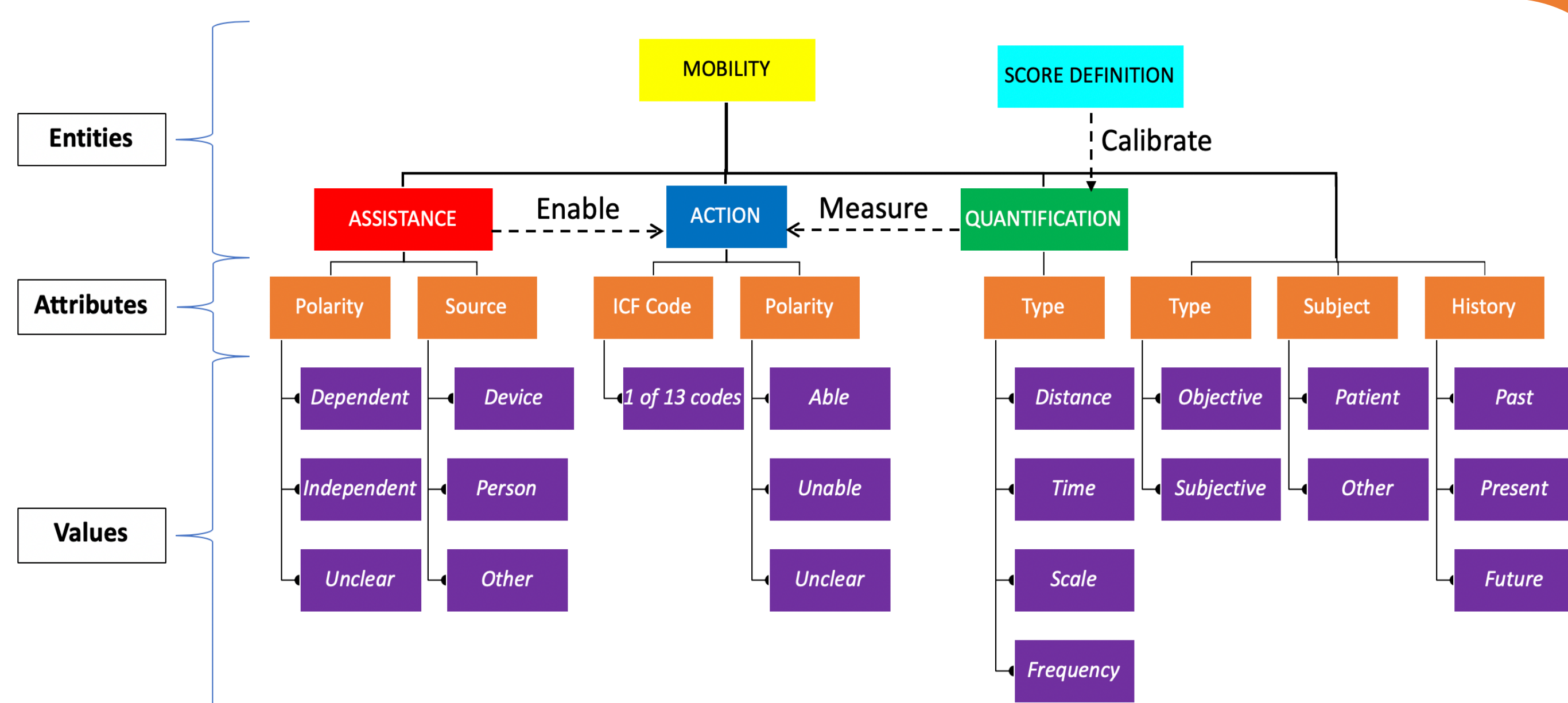


Figure 4: The mobility entity hierarchy comprises of sub-entities, implicit relations, and contextual attributes

Nested Named Entity Recognition (NER)

- Baseline models:
 - Conditional Random Field (CRF): Undirected probabilistic graphical model that computes joint probability of the tag sequence and the original sentence
 - Bidirectional Long-Short Term Memory with CRF decoding: Recurrent neural networks with LSTM cells to retain longer historical information. Substitute softmax with a CRF decoding layer.
- Ensemble learning:
 - Stacking of weak classifiers
 - Error-Correcting Output Code (ECOC)
 - Support Vector Machine (SVM) base classifier

Model	F1-score of exact textual matching				
	Mobility	Action	Assistance	Quantification	Score Definition
CRF	71.26	81.04	68.89	86.92	93.91
Bi-LSTM-CRF	73.04	83.89	71.46	87.95	92.74
Ensemble	75.03	85.55	73.24	88.67	94.05

Table 1: Machine learning model performance

Conclusion

- First comprehensive analysis of an entire domain of the ICF
- First gold standard corpus of mobility functioning information
- State-of-the-art performance on mobility NER

Limitation

- Limited to clinical language at the Clinical Center of the NIH
- Limited to a single domain of the functioning language
- Limited size of the gold standard corpus

¹World Health Organization. International Classification of Functioning, Disability and Health: ICF. 2001

²Cimino, J.J., E.J. Ayres, L. Remennik, S. Rath, R. Freedman, A. Beri, et al., "The National Institutes of Health's Biomedical Translational Research Information System (BTRIS): design, contents, functionality and experience to date". J Biomed Inform, 2014. 52: p. 11-27