# CareCorpus: A Corpus of Real-World Solution-Focused Caregiver Strategies for Personalized Pediatric Rehabilitation Service Design

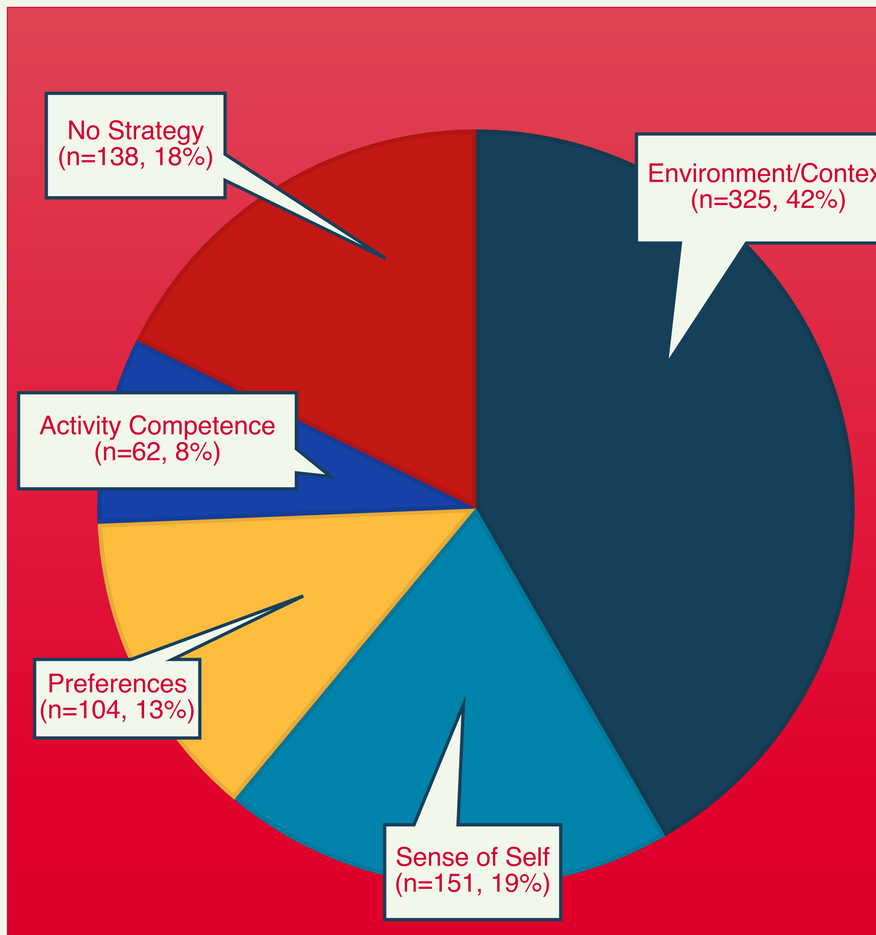
Mina Valizadeh, Vera C. Kaelin, Mary A. Khetani, and Natalie Parde
University of Illinois Chicago
{mvaliz2, vkaeli2, mkhetani, parde}@uic.edu

## Introduction

- Online tools for caregivers of children with rehabilitation needs offer shared strategy banks to support early intervention (EI) service design
- Manually searching through these strategies is time-consuming, especially as strategy banks scale in size
- Automated strategy classification based on real-world clinical frameworks offers potential to greatly reduce caregiver burden in personalized service design

## Data Collection and Annotation

- 780 caregiver strategies sourced from two pilot implementation trials of a large EI program
- Strategies were written by 125 English-speaking caregivers of children ages 0-3 with developmental need who had received EI services for 3+ months
- Caregivers provided strategies through 24 open-ended questions (e.g., “Please describe a strategy that you have tried to help your child participate successfully in basic care routines.”)
- Strategies were dual-annotated and adjudicated by three trained annotators with occupational therapy and/or pediatric rehabilitation expertise
- Annotation categories were known drivers of participation: *environment/context*, *sense of self*, *preferences*, and *activity competence*
- Submitted strategies not belonging to any of those were assigned a “no strategy” label




Distribution of Strategies in CareCorpus

| Class | % Agreement | κ |
|---|---|---|
| Environment/Context | 80.31 | 0.50 |
| Sense of Self | 82.78 | 0.56 |
| Preferences | 72.12 | 0.48 |
| Activity Competence | 75.81 | 0.48 |
| No Strategy | 89.86 | 0.49 |

Per-Class Inter-Annotator Agreement

**Environment / Context**

- Parent is involving toys, setting a time and day and showing the child how to do an activity
- Using pictures, preparing ahead
- Going to the park or taking the child to the park

**Sense of Self**

- Asking the child to help with an activity
- Using positive reinforcement and including the child into an activity
- Encouraging new things, activities, and experiences

**Preferences**

- Including the child's ideas and choosing activities that a child loves and enjoys
- Making activities fun
- Making activities a game and letting the child pick activities or toys

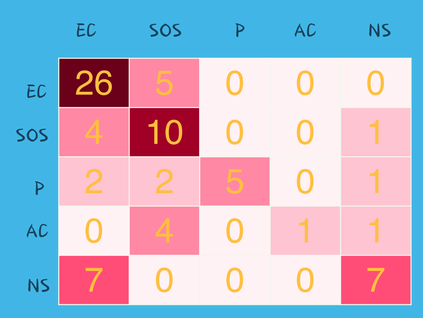
**Activity Competence**

- Working on gross motor and fine motor skills and vision
- Providing instructions about the activity or hand-over-hand support

Common Topics in Strategy Classes

| Model | Accuracy (%) | Precision | Recall | F1 |
|---|---|---|---|---|
| Most Frequent Class | 40.78 | 0.08 | 0.20 | 0.11 |
| Logistic Regression | 57.89 | 0.69 | 0.43 | 0.46 |
| Naïve Bayes | 53.95 | 0.85 | 0.38 | 0.38 |
| **BERT** | **64.47** | **0.73** | **0.53** | **0.56** |
| Bio-ClinicalBERT | 53.94 | 0.71 | 0.40 | 0.39 |

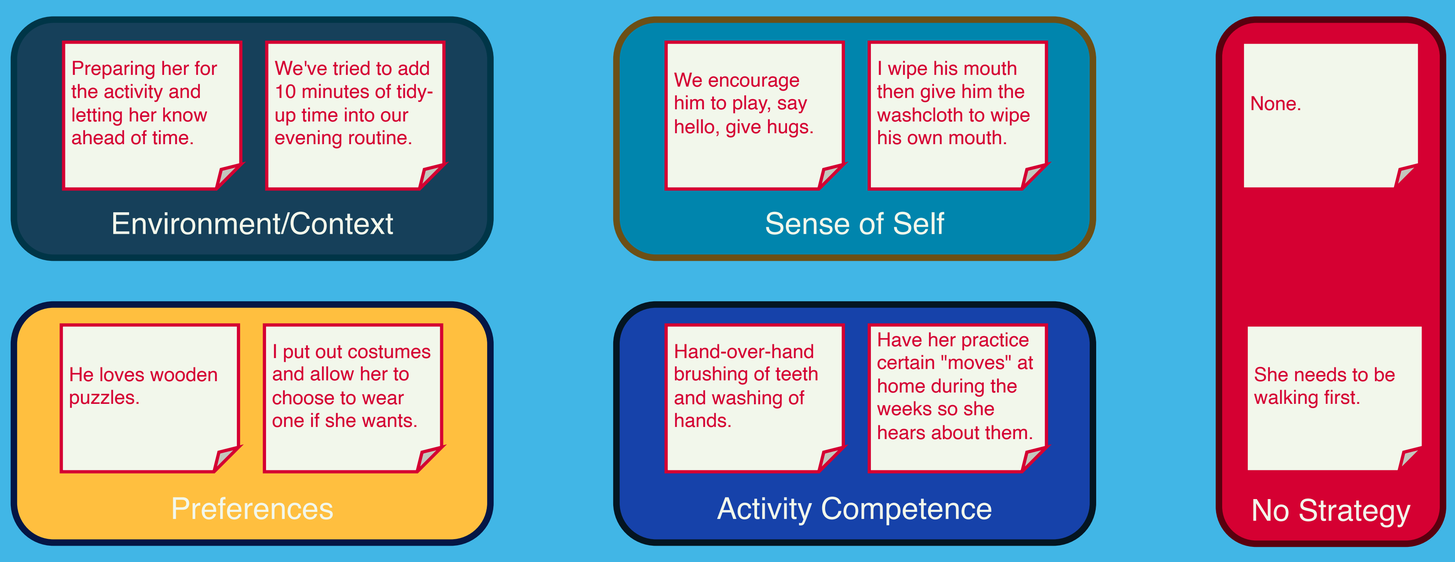
Model Comparison (All-Class Classification)




BERT Predictions

| Model | Accuracy (%) | Precision | Recall | F1 |
|---|---|---|---|---|
| Strategy / No Strategy | 88.15 | 0.82 | 0.76 | 0.79 |
| Extrinsic / Intrinsic | 58.06 | 0.64 | 0.58 | 0.53 |

Subset Classification Performance (BERT)




## Proof of Concept

- Preprocessed strategies through a pipeline that performed spelling correction, punctuation removal, number replacement, case normalization, and (for some models) stopword removal and text lemmatization
- Encoded text using TF-IDF or contextual word embeddings, depending on model
- Experimented with four models for strategy classification: logistic regression, naïve Bayes, BERT, and Bio-ClinicalBERT
- Experimented with five-way classification and pipelined classification strategies

## Key Findings

- **Dataset Validation:** Language models can be fine-tuned to classify caregiver strategies
- **Future Directions:** Task presents rich opportunity to study open-ended, domain-specialized language in a low-resource setting

## Acknowledgements

Data for this study were collected with funding from the National Institutes of Health (NIH 1R03HD084909-01A1 and the American Occupational Therapy Foundation (AOTFIR20KHETANI. Analyses of these data were supported by the National Science Foundation (#2125411), the National Institute on Disability, Independent Living, and Rehabilitation Research (90SFGE0032-01-00), and the University of Illinois Chicago. The content is the responsibility of the authors and does not necessarily represent the official views of these funding agencies.