



# Comparing Grammatical Theories of Code-Mixing

Adithya Pratapa, Monojit Choudhury  
vpratapa@cs.cmu.edu



## 1. Overview

### Two grammatical theories of code-mixing

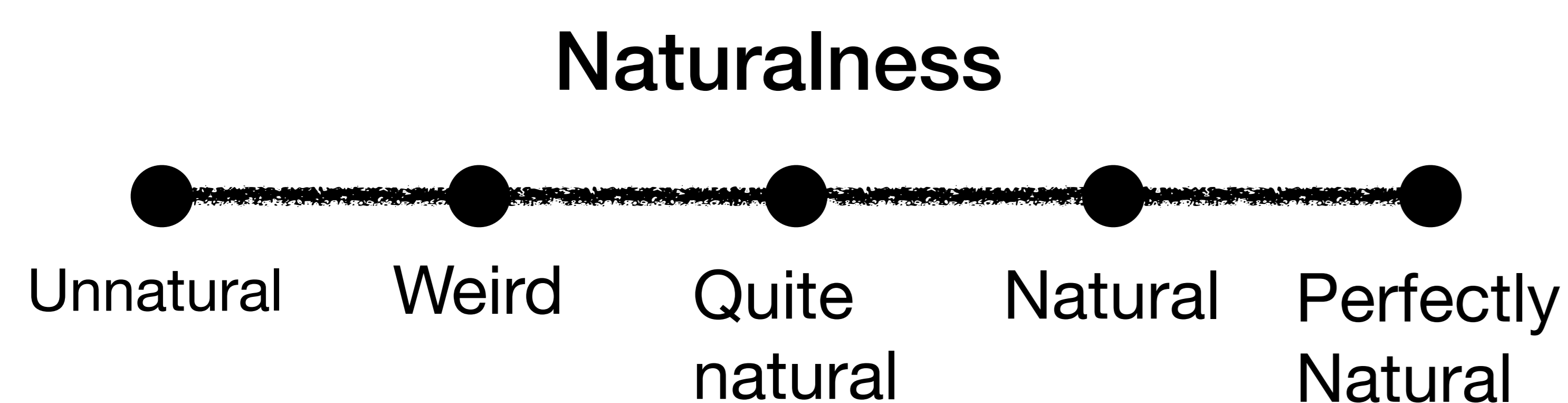
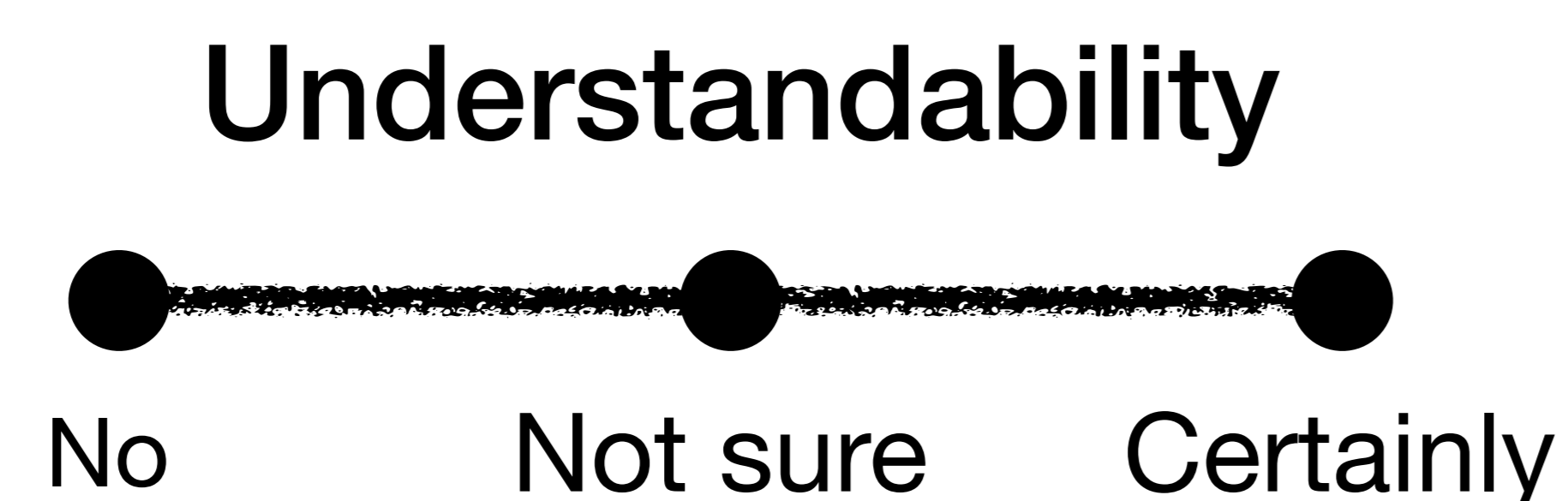
- Equivalence Constraint (Poplack, 1980; EC)
- Matrix-Embedded Language (Myers-Scotton 1993; ML)

### Which grammatical theory results in more natural code-mixed sentences?

- Ask bilingual speakers to rate the sentences on their understandability and naturalness
- Compare against a control group of heuristic models

## 3. Evaluation Setup

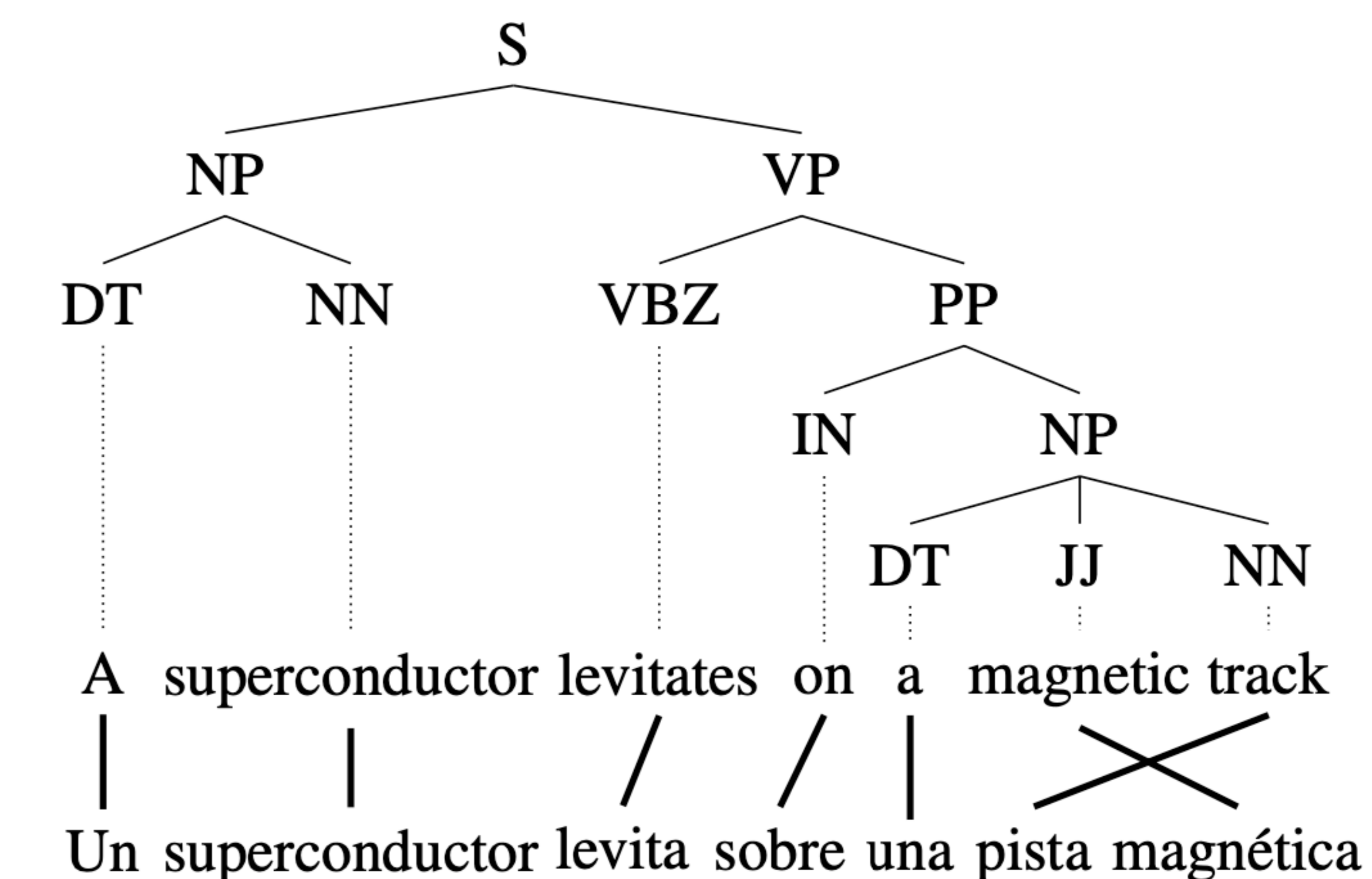
Bilingual Turkers rate code-mixed sentences on these scales



## 2. Generative Models of Code-Mixing

We selectively generate samples from EC and ML theories + 3 heuristic models (aligned, dictionary and parallel)

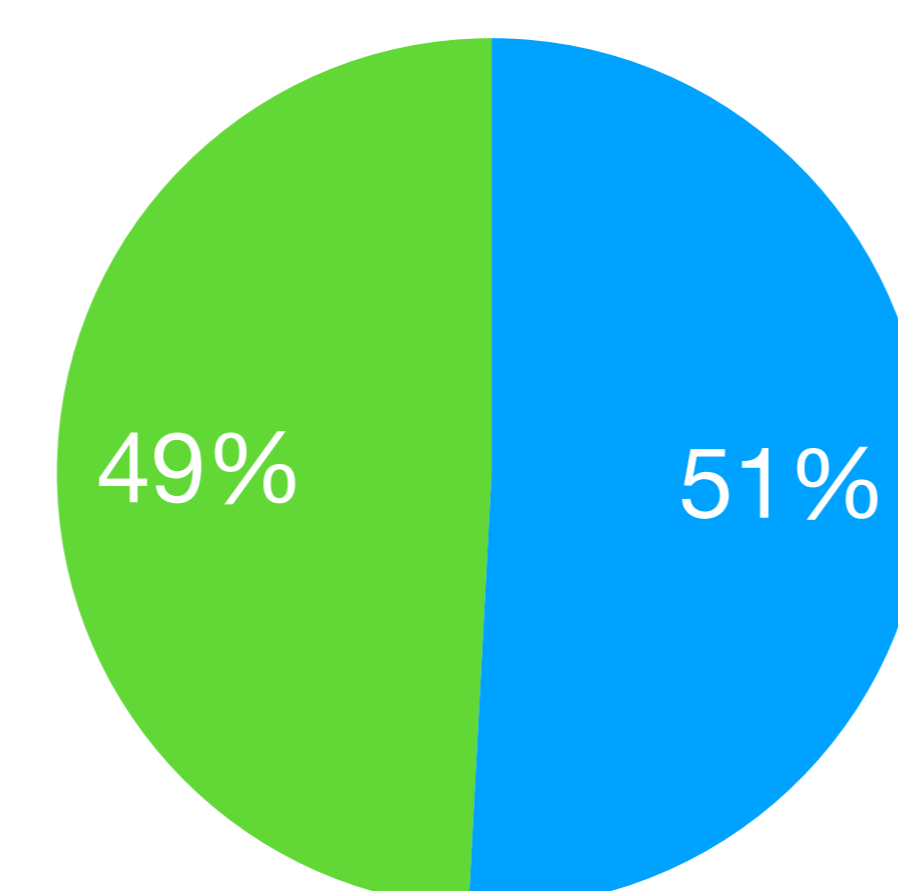
- 1 Un superconductor levita sobre a magnetic track
- 2 Un superconductor levita sobre una pista magnetic
- 3 Un superconductor levita sobre a pista magnética
- 4 A superconductor levitates on a magnetic magnética
- 5 Levita un superconductor sobre a magnetic track



## 4. Results

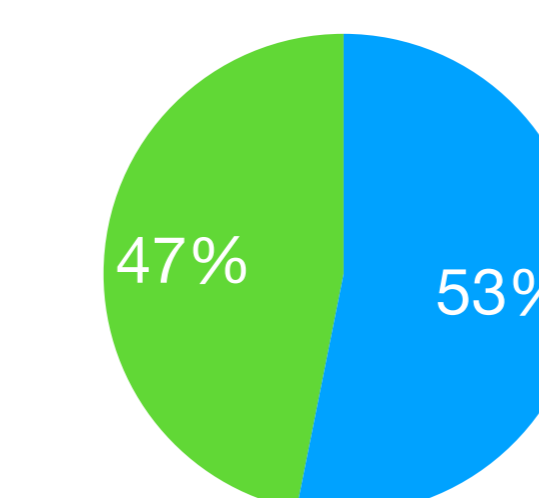
- ML is comparable to EC

ML > EC (blue), ML < EC (green)

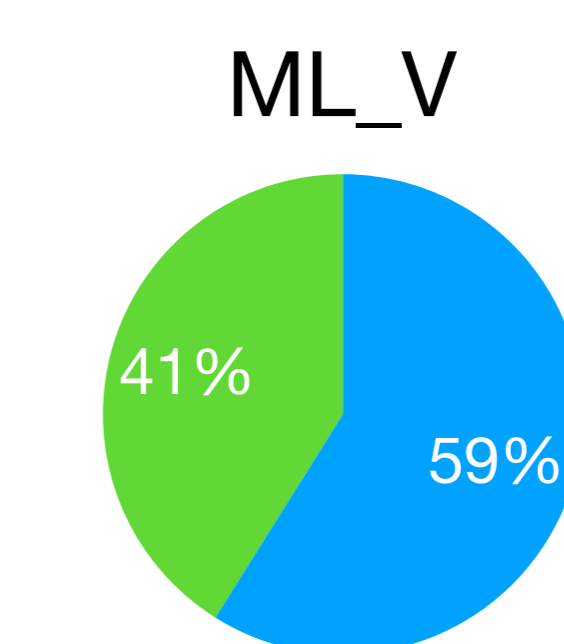
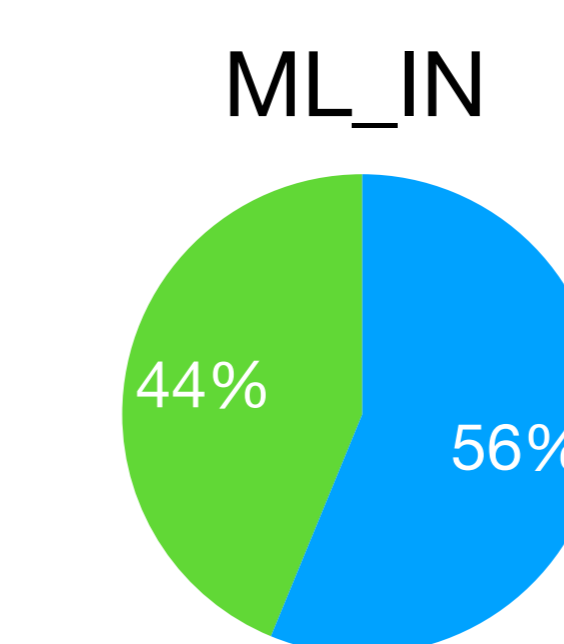
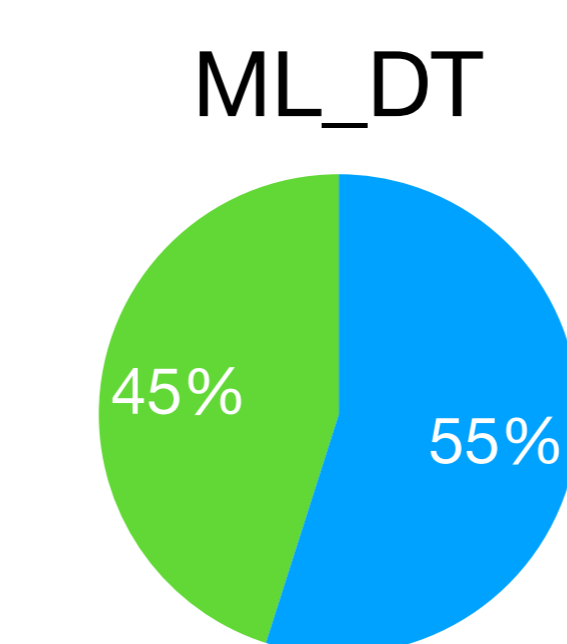


- ML and EC slightly outperform their relaxations

EC > EC\_LS (blue), EC < EC\_LS (green)



M2 / M1



M1 > M2 (blue), M1 < M2 (green)

- ML, EC outperform heuristic models

