

At Your Command!

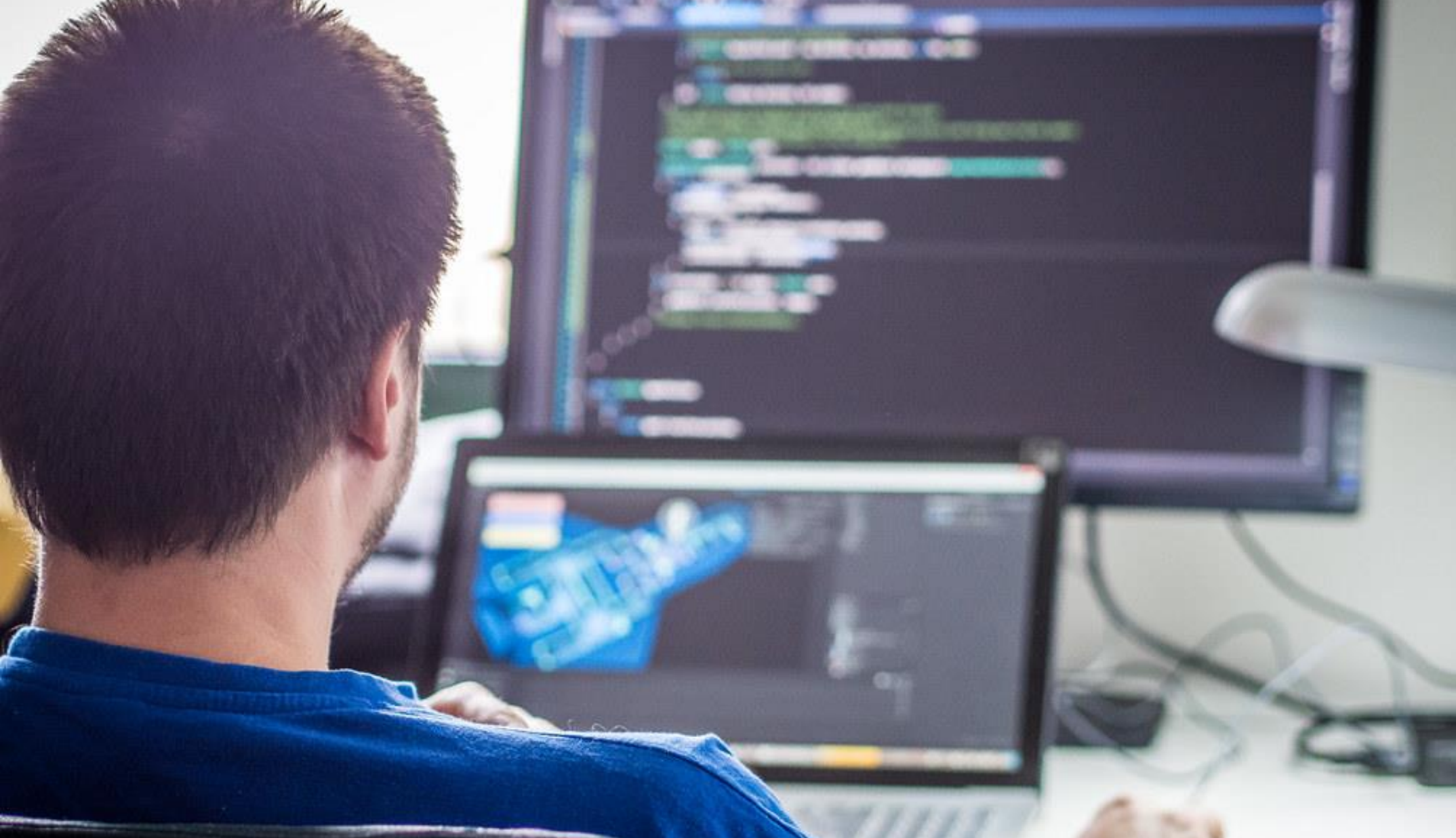
An Empirical Study on How Laypersons Teach Robots New Functions

Sebastian Weigelt, Vanessa Steurer, and Walter F. Tichy

KIT – Department of Informatics – Institute for Program Structures and Data Organization (IPD).









<https://images.squarespace-cdn.com/content/58d2da2a2e69cfb75f606b74/1493080652560-WGJOJ99GBRDFKRM53QP3/Sammet002.jpg?format=1500w&content-type=image%2Fjpeg>

Why is it taking so long?

We lack the empirical foundation!

At Your Command!

An Empirical Study on How Laypersons Teach Robots New Functions

Objective

Collect a dataset of natural language teaching efforts!

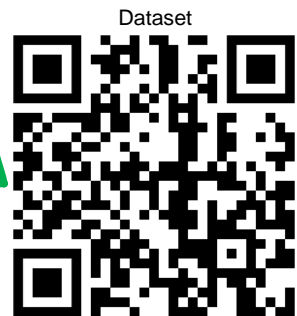
Additional goals:

- reasonable size
- native speakers
- unrestricted language
- unbiased
- domain agnostic
- public available
- ...

Dataset Analysis

- RQ 1** Do laypersons always clearly state that they wish to add new functionality?
- RQ 2** Can the wish for extension be clearly separated from the actions that are to be performed?
- RQ 3** Do laypersons tend to add non-descriptive or meaningless statements?
- RQ 4** Do laypersons repeatedly use particular phrases to declare certain intentions?

<http://dx.doi.org/10.21227/zecn-6c61>



<http://dx.doi.org/10.21227/zecn-6c61>

Related Work

Studies



[1, 2, 3, 3, 4, 5]

Programming in NL



[6, 7, 8, 9, 10]

Corpora



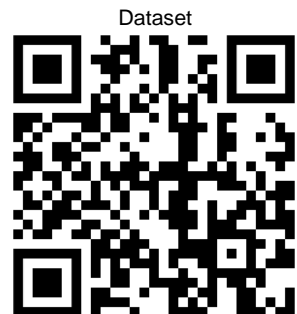
[11, 12, 13]



Scenario-based

Online

Laypeople-oriented



New skill: Start the dishwasher

Intermediate steps:

- The dishwasher can be started by pressing the red button
- It is located in the kitchen
- It needs to be closed before starting it



Some exemplary instructions to teach this new skill:

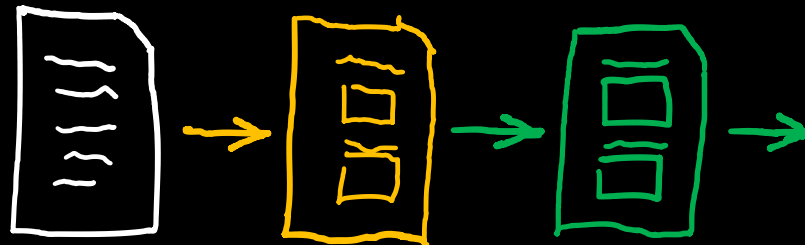
- "Hi Armar, starting the dishwasher means you have to go to the dishwasher, close it and press the red button."
- "You have to close the dishwasher and press its red button for turning it on. That's how you start the dishwasher."

- **Setting**
 - Kitchen
 - Humanoid household robot (ARMAR-III [14])
- Subjects are supposed to teach a new skill, consisting of...
 - an explicitly stated intent to teach a skill
 - intermediate steps to master the skill
- **4 Scenarios (plus a demo scenario)**
 - greeting someone
 - preparing coffee
 - serving drinks
 - setting a table for two
 - (start the dishwasher)

Dataset



Study – Data Collection: Forms



Intro

Example

Scen. 1

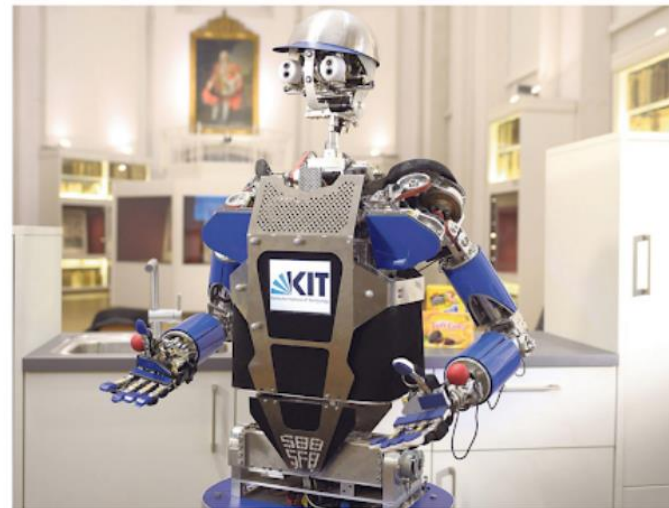
- Goal of Study
- General Setting
- Task

Skill 1: *

New skill: Greet someone

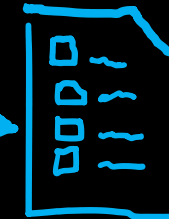
Intermediate steps:

- Look him or her in the eyes
- Wave robot hand
- Say 'Hello'



How would you **teach** the robot this new skill?
"Speak" as naturally as possible. Be creative!

Your instructions:



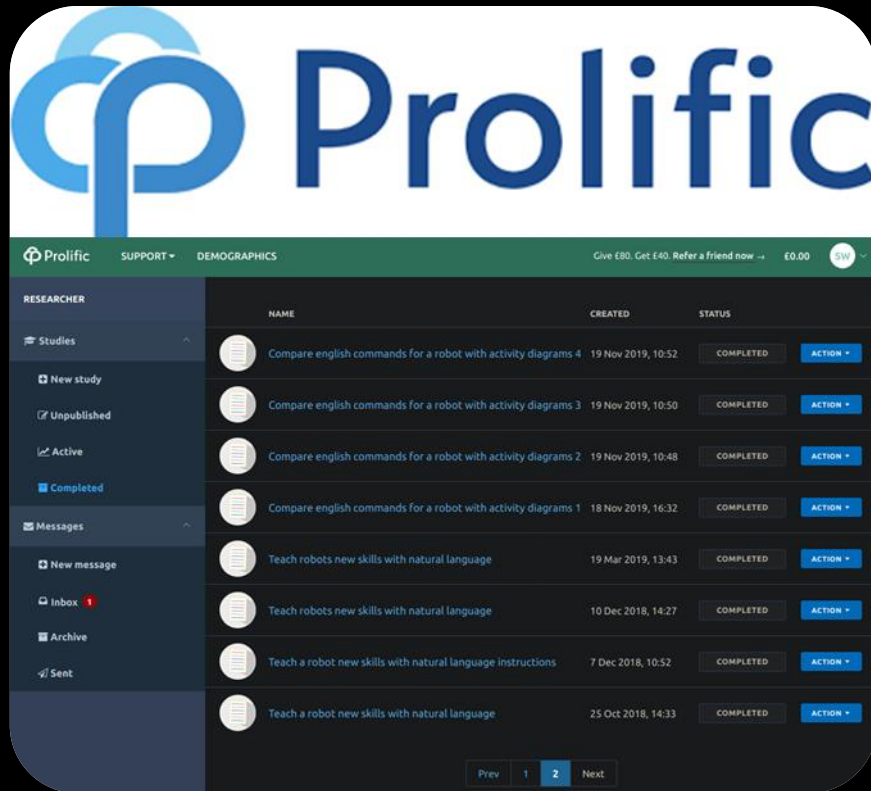
Quest.

- Native Speaker?
- Programming Experience?

Dataset



Study – Data Collection: Data Collection



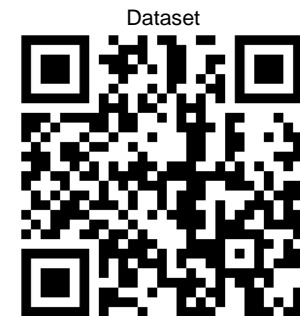
- Platform for online studies
- > 70,000 registered users

Dataset Collection

 870

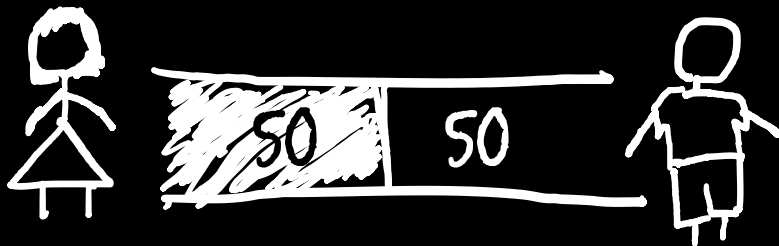
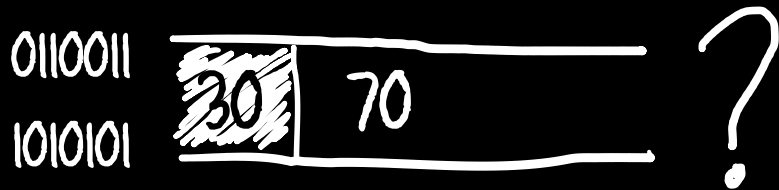
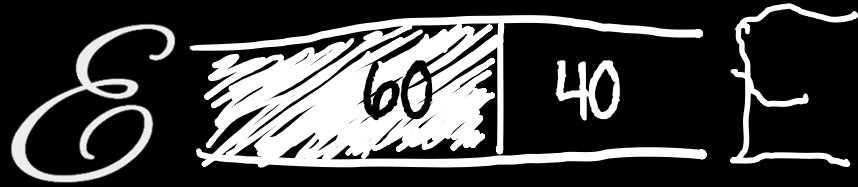
 < 3 days

 3168 (of 3470)



<http://dx.doi.org/10.21227/zecn-6c61>

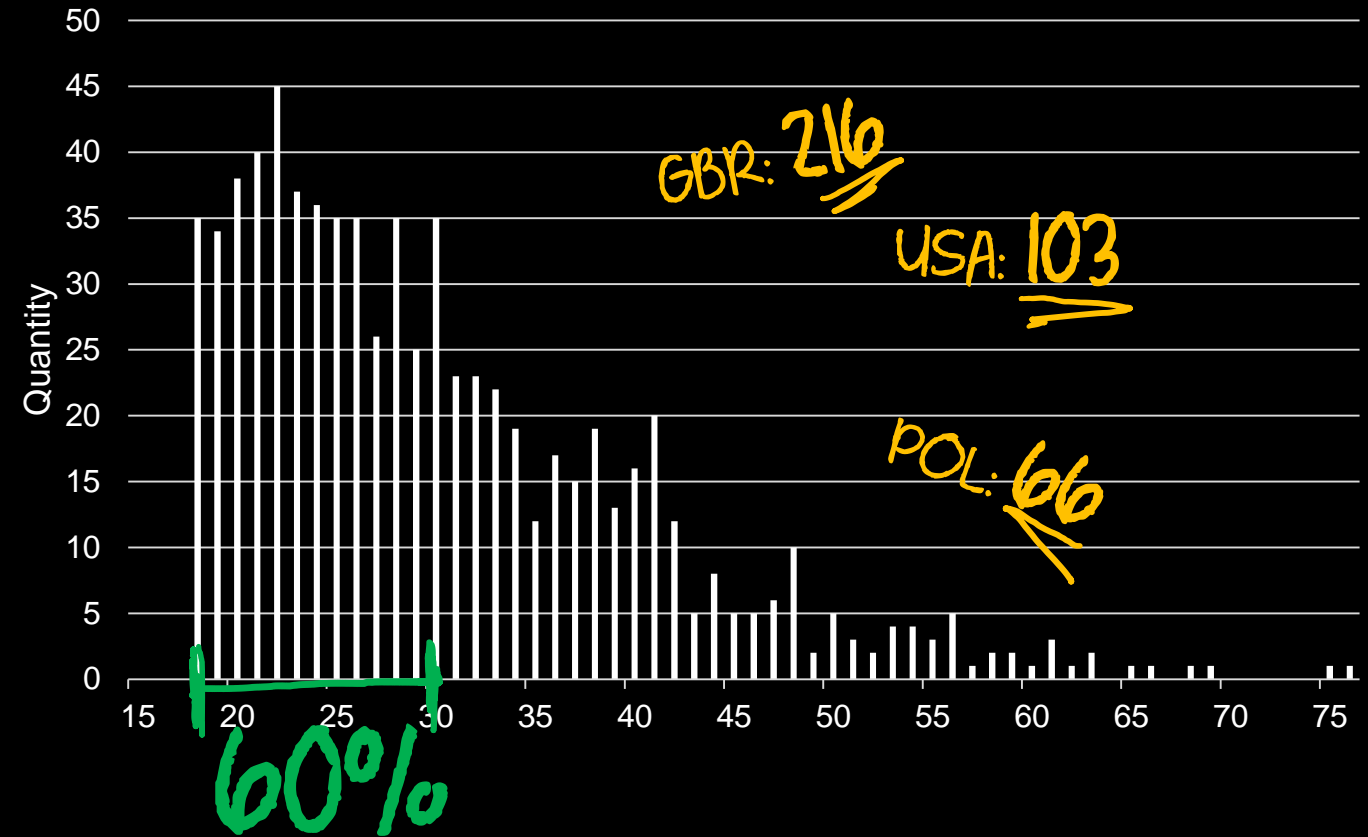
Analysis – Meta Data



minutes to complete



participants' age



Analysis – Dataset: Numbers

words per
submission

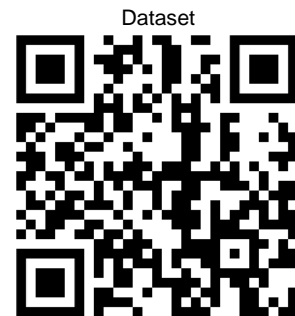
min	1
max	312
avg.	35.43
σ	22.48
sum	109008

unique words
per user

min	2
max	227
avg.	70.69
σ	26.70

unique words
per scenario

<i>greet</i> (S1)	570
<i>coffee</i> (S2)	625
<i>drinks</i> (S3)	691
<i>table</i> (S4)	685
dataset	1469



Analysis – Dataset: Examples

Look directly at the person. Wave your hand. Say 'hello'.

You have to place the cup under the dispenser and press the red button to make coffee.

Making coffee means you have to press the red button, put a cup underneath the hole and then pouring the coffee that comes out into your cup

To bring a beverage, open the fridge and select one of the beverages inside, pour it into one of the glasses on the kitchen counter and hand the glass over to the person.

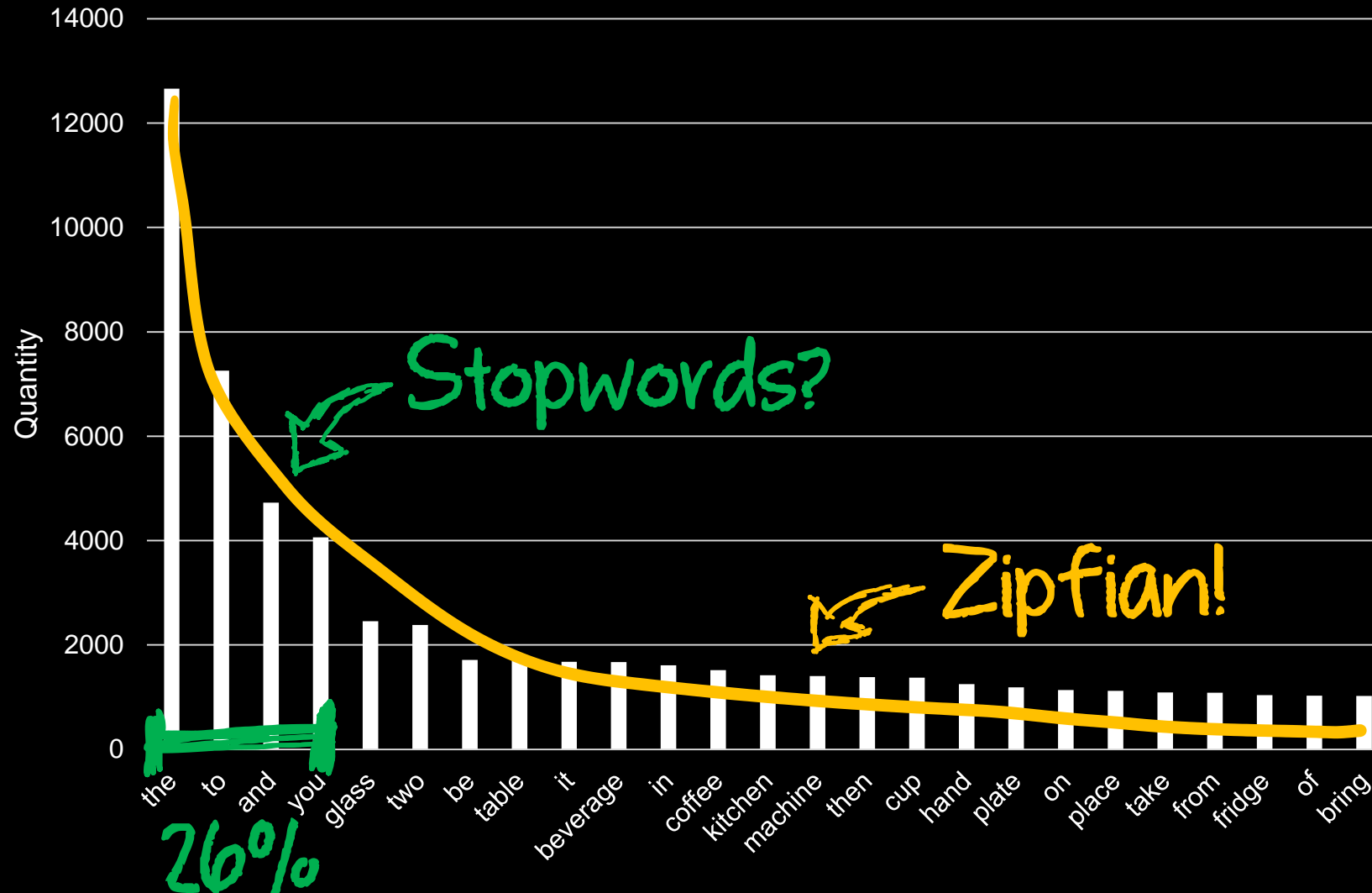
collect cutlery from cupboard, bring them to the table and place down neatly

To set the table for two, Go to the cupboard and take two of each; plates, glasses, knives, and forks. Take them to the kitchen table and set two individual places.



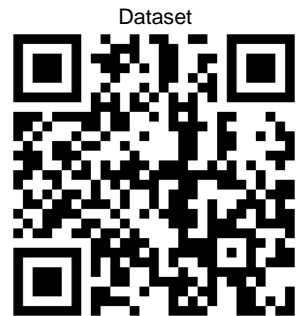
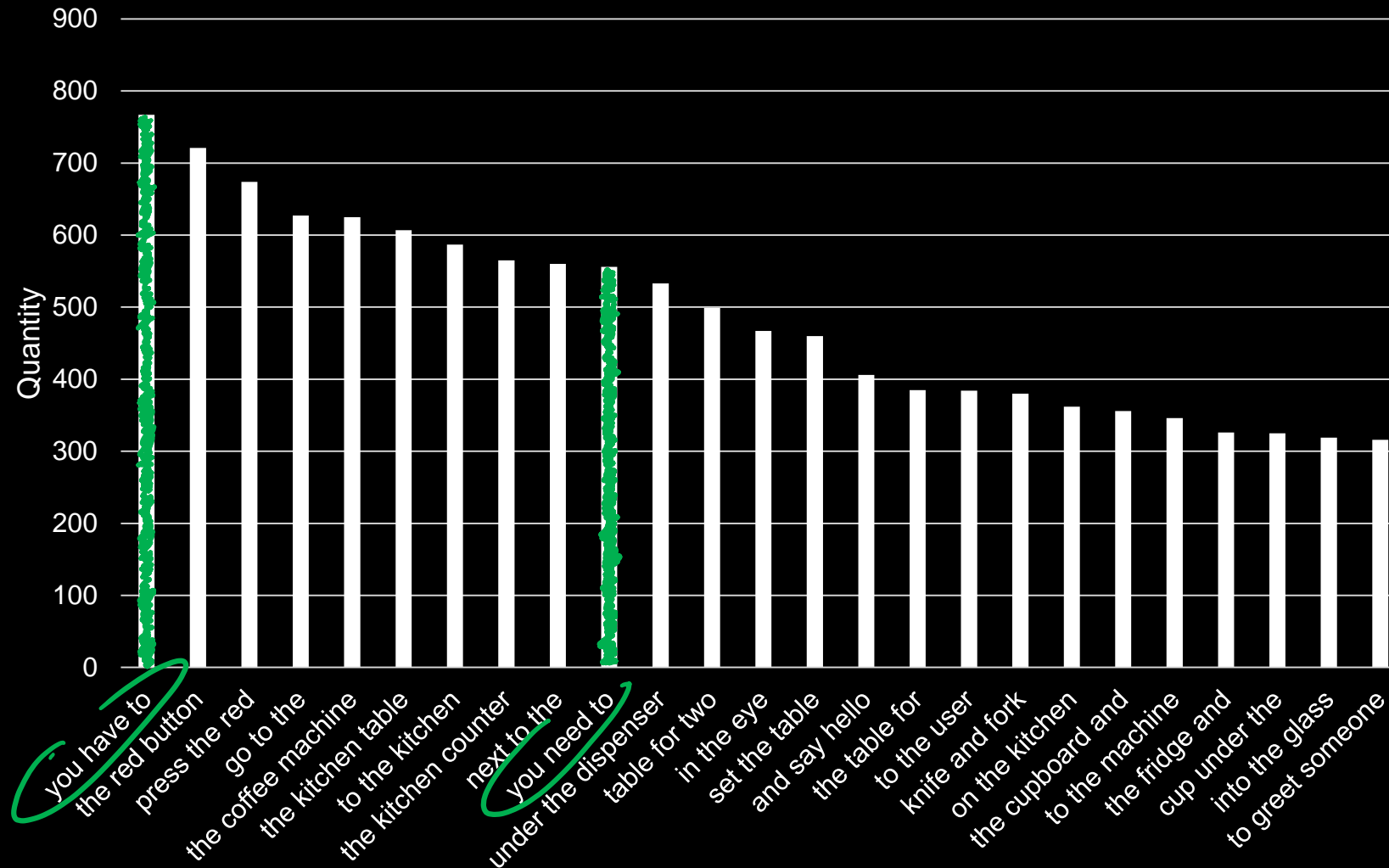
Analysis – Dataset: Most Frequent Words

25 most frequent words



Analysis – Dataset: Most Frequent Trigrams

25 most frequent trigrams



Analysis – Dataset: Qualitative Characteristic

Perspective

Teacher

Developer

Naïve End-User

Structure



Typical Phrases

“we are going to learn how to [...]”

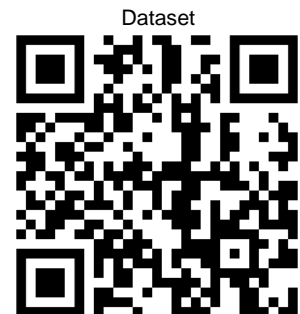
“in order to [...] you have to [...]”

“if you want to [...] you need to [...]”

“[...] means you have to [...]”

Grammar

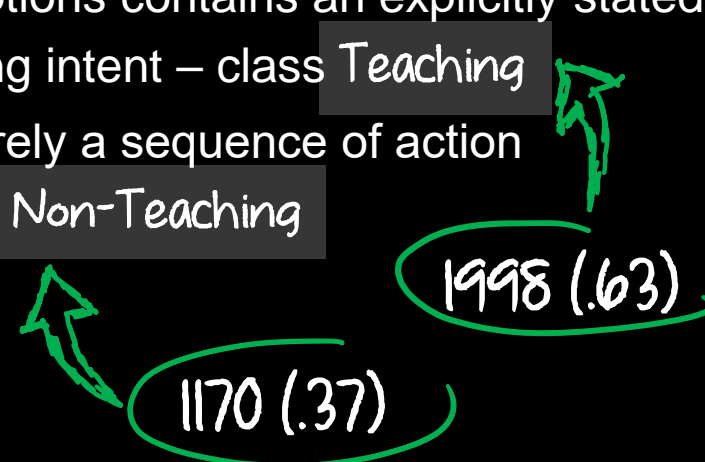
most	some	a few
<ul style="list-style-type: none">• present tense• imperative	<ul style="list-style-type: none">• declarative• interrogative clauses	<ul style="list-style-type: none">• passive voice• conjunctive mood



Labeling

Teaching effort (binary):

- descriptions contains an explicitly stated teaching intent – class **Teaching**
- it's merely a sequence of action – class **Non-Teaching**



Semantic structure (ternary):

Phrases of teaching efforts either...

- 15,559 (.21) } **declare** the new function (wish for extension and a name), or...
- 57,156 (.76) } **specify** the intermediate steps of the function to be learned, or...
- 2,219 (.03) } have **miscellaneous** content (irrelevant in our context)

hey Robo preparing a cup of coffee means you have to put a coffee mug under the dispenser and then press the red button on the coffee machine that's how you make some coffee

Teaching

collect cutlery from cupboard, bring them to the table and place down neatly

Non-Teaching



Research Questions (Revisited)

No, unfortunately not!
(.37 of all descriptions without teaching intent)

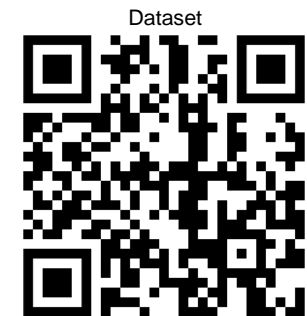
RQ 1 Do laypersons always clearly state that they wish to add new functionality?

RQ 2 Can the wish for extension be clearly separated from the actions that are to be performed?

RQ 3 Do laypersons tend to add non-descriptive or meaningless statements?

RQ 4 Do laypersons repeatedly use particular phrases to declare certain intentions?

Unclear! There are some trends
(but nothing to lean on)



90% of the submission
were **rejected**

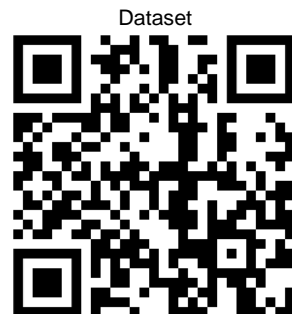
Christensen: “[...] this approach is **reasonable**, since this kind of study aims to analyze the participants verbal knowledge transfer and **not their willingness to achieve good results** or their interpretation of the exercises” [15]

spoken vs. written
language

- encouraged to respond spontaneously
- unrestricted language
- still **no “real” spoken language**


all scenarios have the
same basic setting

- subjects might act **differently** in **other domains**
- some findings are domain agnostic



Conclusion & Future Work

Objective: Collect and analyze a dataset of natural language teaching efforts

- online study on  Prolific
- setting: robot in a kitchen
- subjects gave descriptions to 4 scenarios

Data Analysis

- only two-thirds have teaching intents
- teaching intents in m.f. trigrams
- no uniform structure

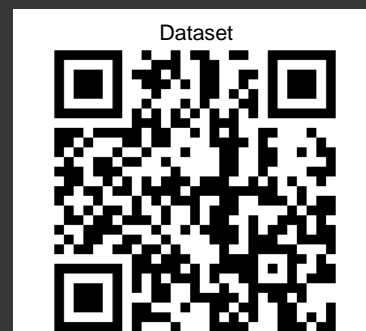
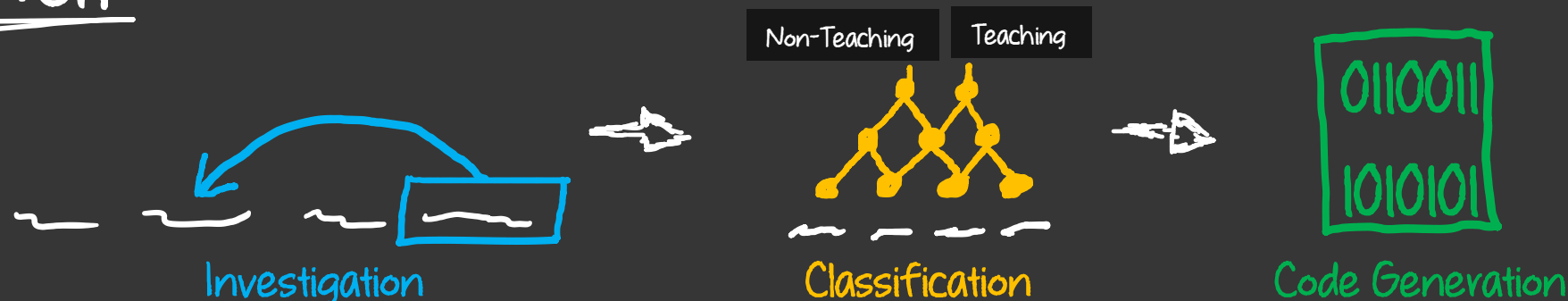


870



3168

Future Work



Dataset

<http://dx.doi.org/10.21227/zecn-6c61>

References (1)

- [1] L. Miller, “Natural language programming: Styles, strategies, and contrasts,” IBM Systems Journal, vol. 20, no. 2, pp. 184–215, 1981.
- [2] J. Nielsen, Usability Engineering. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.
- [3] A. W. Biermann, B. W. Ballard, and A. H. Sigmon, “An Experimental Study of Natural Language Programming,” International Journal of Man-Machine Studies, vol. 18, no. 1, pp. 71–87, 1983.
- [4] J. F. Pane, C. Ratanamahatana, and B. A. Myers, “Studying the Language and Structure in Non-programmers’ Solutions to Programming Problems,” International Journal of Human-Computer Studies, vol. 54, no. 2, pp. 237–264, 2001.
- [5] H. Lieberman and H. Liu, “Feasibility Studies for Programming in Natural Language,” in End User Development, ser. Human-Computer Interaction Series, H. Lieberman, F. Paterno, and V. Wulf, Eds. Springer Netherlands, Jan. 2006, no. 9, pp. 459–473.
- [6] V. Le, S. Gulwani, and Z. Su, “SmartSynth: Synthesizing smartphone automation scripts from natural language,” in Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services - MobiSys ’13. Taipei, Taiwan: ACM Press, 2013, p. 193.
- [7] L. She, Y. Cheng, J. Y. Chai, Y. Jia, S. Yang, and N. Xi, “Teaching Robots New Actions through Natural Language Instructions,” in The 23rd IEEE International Symposium on Robot and Human Interactive Communication. Edinburgh, UK: IEEE, Aug. 2014, pp. 868–873.

References (2)

- [8] I. Markievicz, M. Tamosiunaite, D. Vitkute-Adzgauskiene, J. KapociuteDzikiene, R. Valteryte, and T. Krilavicius, “Reading Comprehension of Natural Language Instructions by Robots,” in *Beyond Databases, Architectures and Structures. Towards Efficient Solutions for Data Analysis and Knowledge Representation*. Springer, May 2017, pp. 288–301.
- [9] M. Manshadi, D. Gildea, and J. Allen, “Integrating Programming by Example and Natural Language Programming,” in *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, ser. AAAI’13. AAAI Press, 2013, pp. 661–667.
- [10] T. J. Li, I. Labutov, X. N. Li, X. Zhang, W. Shi, W. Ding, T. M. Mitchell, and B. A. Myers, “APPINITE: A Multi-Modal Interface for Specifying Data Descriptions in Programming by Demonstration Using Natural Language Instructions,” in *2018 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, Oct. 2018, pp. 105–114.
- [11] S. Tellex, T. Kollar, S. Dickerson, M. R. Walter, A. G. Banerjee, S. J. Teller, and N. Roy, “Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation.” in *AAAI*, 2011.
- [12] E. Bastianelli, G. Castellucci, D. Croce, L. Iocchi, R. Basili, and D. Nardi, “HuRIC: A human robot interaction corpus,” in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*, 2014, pp. 4519–4526.

References (2)

- [13] J. Sales, S. Handschuh, and A. Freitas, “SemEval-2017 Task 11: EndUser Development using Natural Language,” in Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada: Association for Computational Linguistics, Aug. 2017, pp. 556–564.
- [14] T. Asfour, K. Regenstein, P. Azad, J. Schroder, A. Bierbaum, N. Vahrenkamp, and R. Dillmann, “ARMAR-III: An integrated humanoid platform for sensory-motor control,” in Humanoid Robots, 2006 6th IEEE-RAS International Conference On. IEEE, 2006, pp. 169–175