# Benchmarking Pretrained Language Models for Italian Natural Language Understanding

Anonymous ACL submission

## Abstract

Since the advent of Transformer-based, pretrained language models (LM) such as BERT, Natural Language Understanding (NLU) components in the form of Dialogue Act Recognition (DAR) and Slot Recognition (SR) for dialogue systems have become both more accurate and easier to create for specific application domains. Unsurprisingly however, much of this progress has been limited to the English language due to the existence of very large datasets in both dialogue and written form. In this paper, we use the newly released JILDA dataset to benchmark three of the most recent pretrained LMs: Italian BERT, Multilingual BERT, and AlBERTo. Results show that the monolingual version of BERT performs better than both the multilingual one and AlBERTo. This paper highlights the challenges that still remain in creating effective NLU components for lower resource languages, and constitutes a first step in improving NLU for Italian dialogue.

## 1 Introduction

The field of Natural Language Processing was transformed when Vaswani et al. (2017) presented their self-attention-based, Transformer model for representation or embedding of Natural Language (NL) strings, with Devlin et al. (2019) then releasing BERT, a large scale pretrained LM, showing that new state of the art results could be obtained in many canonical NLP tasks just by fine-tuning with one additional task-specific output layer. This *transfer learning* methodology has also been applied to our problem of interest in this paper: that of Dialogue Act Recognition (DAR, e.g. Chakravarty et al. (2019)) combined with Slot Recognition (SR), forming the basis of the most important component in dialogue systems (henceforth DS) today: Natural Language Understanding (NLU). Much of the progress above has, however, been limited to the English language due largely to the unavailability of high quantities of language corpora in other languages. In comparison to English, in which there are numerous dialogue datasets available (see Lowe et al. (2017); Li et al. (2018); Budzianowski et al. (2018); Liu et al. (2021) among many others), Italian is a lower-resource language and, with few exceptions (Mana et al., 2004; Castellucci et al., 2019), there is currently a paucity of dialogue datasets available with appropriate Dialogue Act & Slot annotations for training effective NLU models. Large scale multilingual models do exist (e.g. Multilingual BERT), but it is as yet unclear how these models *transfer* to the NLU tasks of DAR & SR. One important reason for this uncertainty is that nearly all existing, large-scale LMs have been trained on open domain, written language, whereas dialogue is known to be very different from text or written language: dialogue is highly context-dependent, is replete with fragments (Fernández and Ginzburg, 2002; Purver et al., 2009), ellipsis (Colman et al., 2008) & disfluencies (Shriberg, 1996; Hough, 2015), and is highly domain-specific (Eshghi et al., 2017). Noble and Maraev (2021) provide evidence for this, showing that pretrained BERT does not transfer well for the DAR task without being fine-tuned on the target dialogues. In this paper, we focus on NLU for dialogue systems in Italian. We use the newly released JILDA corpus (Sucameli et al., 2020) – one of the very few Italian dialogue datasets in the public domain – to evaluate three of the most recent pretrained LMs on the DAR & SR tasks: Multilingual BERT (Devlin et al., 2019), Italian BERT (Schweter, 2020), and AlBERTo (Polignano et al., 2019).

## 2 Related work

Ever since the advent of the Transformer model, BERT (Devlin et al., 2019) has become the de facto standard for the DAR and SR tasks, and has seen success in many dialogue domains in the English language (Mehri et al., 2019; Ribeiro et al., 2019; Chakravarty et al., 2019; Bao et al., 2020). For these tasks, a *transfer learning* method is employed using BERT, which uses a multi-layer bidirectional transformer to embed the input text. In such approaches, BERT is used as the pre-trained encoder, whose one or more hidden layers are fed to additional output layer(s) or classifiers and fine-tuned on specific in-domain NLU datasets. Considering the effectiveness of such a transfer learning approach for dialogue, Noble and Maraev (2021) show, interestingly, that the pretrained model isn't of much use without fine-tuning on target dialogue data. In this paper, we study the usefulness of three different versions of BERT as the pretrained language model, and evaluate their performance in the DAR & SR tasks on the JILDA dataset, a collection of mixed-initiative, human-human dialogues in Italian, and in the 'job offer' domain (Sucameli et al., 2020). JILDA consists of 745 dialogues, 17,889 utterances, and a total of 263,104 tokens, and it is characterised by great linguistic variability and lexical complexity.

## 3 Models

Our experiments were conducted within ConvLab-2 (Zhu et al., 2020): an open-source multi-domain end-to-end dialogue system platform. For our experiments we decided to use **BERTNLU**, a ConvLab-2 NLU multi-task module based on a pretrained BERT to which it adds on top two Multi-Layer Perceptrons (MLPs), one for intent classification and another for slot tagging, as shown in Figure 1. Here, the Transformer model is called at different times within the same cycle. The number of layers depends on the pretrained LM used. For each sentence, it is called twice with the indicated inputs and outputs, and also produces a pooled representation of the context. Then, the Slot Classifier produces as many outputs as the words in the sentence, while the DAR returns a score on the different DA values. In BERTNLU all those dialogue acts which appear in the utterances are converted using BIO tags, a common tagging format for tagging tokens in chunks (Ramshaw and Marcus, 1995).

| | bert-italian-xxl | bert-multil. | AlBERTo |
|---|---|---|---|
| **Voc. Size** | 32K | 119K | 128K |
| **Source** | OPUS, OSCAR and Wikipedia | Wikipedia | TWITA |

Table 1: Comparison of vocabulary size of the LMs

We used BERTNLU combined with three different language models available on Hugging Face: **bert-base-italian-xxl-cased**[1] (Schweter, 2020), **bert-multilingual-cased**[2] (Devlin et al., 2019) and **AlBERTo**[3](Polignano et al., 2019). The first one is trained on Wikipedia, the OPUS corpus and the Italian part of the OSCAR corpus. The second one is trained with the top 100 languages from Wikipedia, including Italian. Since the size of Wikipedia varies from language to language, and to avoid under-representation of low resource languages, in the multilingual version of BERT, high-resource languages (like English) are under-sampled, while low-resource languages are over-sampled. Finally **AlBERTo** (Polignano et al., 2019) is a BERT LM for the Italian language, trained on 200M tweets with a vocabulary size of 128k. AlBERTo replicates the BERT stack and it is trained using masked language modelling loss only.

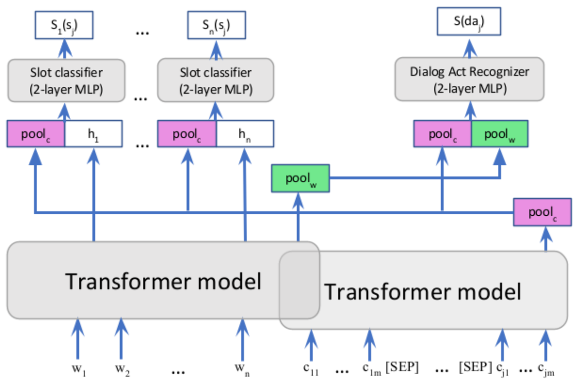


Figure 1: BERTNLU architecture. The Transformer models produce two types of pools, one for the words (w) and another for the contexts (c). These pools are sent to the Slot Classifier and the Dialogue Act Recognizer. There are as many Slot Classifiers as there are words, while for the Dialogue Act is produced a single distribution of probability on the different values.

## 4 Experiments

We use the JILDA dataset to finetune & evaluate the above-mentioned models on the DAR & SR tasks. We use 80% of the data for training (596 dialogues) & 20% for testing and validation (respectively, 75 and 74 dialogues). The hyper-parameter tuning procedure is described in Appendix 7.1. After fixing the hyper-parameters, we trained each model and computed average scores for Precision, Recall and F1 Score. In order to quantify how

[1] https://github.com/dbmdz/berts
[2] https://github.com/google-research/bert
[3] https://github.com/marcopoli/AlBERTo-it

well each pretrained encoder – bert-base-italian, bert-multilingual & AlBERTo – encodes the target JILDA dialogues, i.e. how well it transfers, we evaluated each model in two training conditions: (1) **end-to-end**, where the weights of the underlying encoder model were finetuned together with the task-specific DAR & SR layers; and, (2) **frozen-lm** where all the weights of the encoder layers were frozen during training with only the task-specific layers fine-tuned.

## 5 Results & Discussion

**The end-to-end condition** Table 2 shows the averaged results obtained for the three models in the end-to-end condition. The overall results record those cases in which both the DAs and the slots in a sentence have been correctly predicted.

| | | bert-ita-xxl | bert-multi | AlBERTo |
|---|---|---|---|---|
| **Acts** | Prec. | 81.55 | **82.85** | 79.74 |
| | Rec. | **75.36** | 70.41 | 70.66 |
| | F1 | **78.33** | 76.12 | 74.92 |
| **Slots** | Prec. | **71.65** | 68.06 | 70.78 |
| | Rec. | **71.27** | 66.99 | 65.60 |
| | F1 | **71.46** | 67.52 | 68.09 |
| **Overall** | Prec. | **74.20** | 71.66 | 73.13 |
| | Rec. | **72.38** | 67.92 | 66.97 |
| | F1 | **73.28** | 69.74 | 69.92 |

Table 2: Values of Precision, Recall and F1 Scores in the `end-to-end` condition.

Analysing the performance reported in Table 2, the best performing model definitely appears to be **bert-ita-xxl**. Comparing the monolingual models (bert-ita-xxl vs. AlBERTo) we noticed that bert-ita shows a superior performance than AlBERTo, which, however, has a larger vocabulary than the first one (see Tab. 1). This result is probably due to the fact that the original training dataset of bert-ita includes transcripts of spoken conversation and subtitles, which present a syntactic and semantic structure close to the one of the JILDA dialogues. On the other hand, AlBERTo is trained on Italian tweets, which tend to have a simplified structure compared to that of the dialogues. In addition to this, we observed that the difference in performance between the multi-lingual and monolingual BERT models is small, and that the multilingual BERT model is therefore not less effective. This shows that at least the Italian language is represented well within the multilingual BERT model.

The results achieved are good if we consider that they were obtained using extremely complex training data. Table 3 compares the results achieved by JILDA with bert-ita-xxl, our best model, with those obtained by MultiWOZ 2.1 (Eric et al., 2020) and reported in (Han et al., 2021), where the dataset is used to train, via ConvLab, the BERTNLU module for the DAR and SR tasks[4]. Although the F1 scores gained with JILDA are inferior to those obtained with MultiWOZ, they seem to be not only reasonable but also very positive, since our model was trained using a dataset which is much smaller (JILDA has 745 dialogues and 263K tokens, while MultiWOZ includes over 10K dialogues and 1M tokens) and, at the same time, much richer from a lexical point of view. In fact, the number of values extracted from the lexical vocabulary of each slot is 5.779 in JILDA and 2.111 in MultiWOZ.

| Datasets | F1 (Slot/DA/Both) |
|---|---|
| JILDA | 71.46/78.33/73.28 |
| MultiWOZ 2.1 | 81.18/88.34/83.77 |

Table 3: Performance of BERTNLU with JILDA and MultiWOZ 2.1.

Taking into account all these considerations, it seems that the NLU model trained on JILDA presents convincing and competitive results.

**The frozen-lm condition** Table 4 shows the averaged Precision, Recall & F1 Score values obtained in the `frozen-lm` condition where the weights of the encoder stack were frozen during training and only the task-specific heads fine-tuned.

| | | bert-ita-xxl | bert-multi | AlBERTo |
|---|---|---|---|---|
| **Acts** | Prec. | 82.26 | **96.00** | 80.13 |
| | Rec. | 32.01 | 10.57 | **54.51** |
| | F1 | 46.09 | 19.05 | **64.66** |
| **Slots** | Prec. | 70.15 | 63.80 | **72.23** |
| | Rec. | **55.34** | 48.26 | 50.22 |
| | F1 | **61.87** | 54.96 | 59.25 |
| **Overall** | Prec. | 72.02 | 65.44 | **74.34** |
| | Rec. | 49.05 | 38.10 | **51.38** |
| | F1 | 58.35 | 48.16 | **60.77** |

Table 4: Values of Precision, Recall and F1 Score recorded for the three models without fine-tuning the language model encoder stack.

Comparing Table 2, which shows the performance of the fine-tuned models, with Table 4, it is clear that the presence of fine-tuning allows to

[4]For MultiWOZ 2.0 no data relating to NLU training is reported, thus we compared our results with the directly following version of the dataset.

gain better values. The results above are in line with those found by (Noble and Maraev, 2021) and highlight the importance of fine-tuning pre-trained encoders. Interestingly however, comparing the performance of the three models, when the fine-tune parameter is set to false, the one which performs better is AlBERTo. We believe that this is due to the data and vocabulary size used in the original training; in fact, AlBERTo presents 191GB of raw data and a vocabulary of 128K terms, while bert-ita consists of 81GB of data and 32K terms. In the absence of fine-tuning it seems that AlBERTo is it able to obtain better performances.

**Error Analysis** Having computed the F1 scores of the three models, we conducted an error analysis in order to verify which acts and slots were recognised more easily and which with more difficulties. To this end, we calculated the accuracy for DA and slot and for each of the models. This measure is often used to evaluate NLU models and for intent detection task (Mohamad Suhaili et al., 2021), which is similar to our DAR and SR tasks.

| | bert-ita-xxl | bert-multi | AlBERTo |
|---|---|---|---|
| DA Acc. | **78.25** | 76.03 | 74.84 |
| Slot Acc. | **71.46** | 67.57 | 68.08 |

Table 5: Averaged accuracy in DAR and SR tasks

As shown in Table 5, the accuracy values obtained are positives, especially for the DAR task. Analysing the accuracy of each DA, we noticed that *inform* had the highest values, while *greet* the lowest, probably due to the number of representation in the dataset of these acts (see the Appendix for the number of DAs occurrences in JILDA ).

Regarding the classification of slots, it seems that the models have more difficulty with those slots which share lexical entries. For instance, the label relating to the *area* slot is frequently marked with *degree* while *job-description* is often marked as *duties*. This probably happens because those slots tend to occur in the same linguistic contexts and to share part of their lexical vocabularies. For example, in Fig. 2 the text span can be annotated both with the slot *area* and with *degree*, due to their vocabulary overlap. The analysis and the discussion conducted, point out that creating effective NLU components for dialogue systems in domains grounded in data as linguistically rich & complex as JILDA remains a challenge. Therefore, starting from the values presented in Tab. 2, we propose in the future to further investigate the DAR and SR tasks for NLU Italian models, training the models in order to achieve even a better performance.

**I am looking for a job in my field of study. I graduated in Economics and marketing in Turin.**

| | | |
|---|---|---|
| True label | area | Economics and marketing |
| Predicted label | **degree** | Economics and marketing |

Figure 2: Overlap of slots' lexical vocabularies

## 6 Conclusion

In this paper we have evaluated three of the most recent pretrained LMs, namely Italian BERT, Multilingual BERT and AlBERTo, on JILDA , a newly released corpus of Italian dialogues in the job application domain. We fine-tuned and tested these models on the Dialogue Act Recognition and Slot Recognition tasks which are good proxy tasks for how well and under what training conditions these models are able to effectively encode dialogue semantics. Our results showed that: (1) comparing the monolingual and the multilingual models, the first type resulted to be more able to obtain a better performance when trained on an Italian dialogic dataset; (2) the size of the dataset used in the original training of the LM has less impact on the results than the type of data used in the original training; in fact, it was recorded a better performance for bert-ita-xxl, whose vocabulary is smaller than the one of AlBERTo but includes data which have linguistic features close to those of the JILDA dialogues, respect than the model pre-trained with a large collection of tweets; (3) the multilingual BERT model performs only slightly worse than the monolingual model, highlighting the relative effectiveness of the multilingual model for the Italian language; and (4) fine-tuning the pretrained encoder is important, especially when the target data are dialogues that differ in many important ways from written data. Furthermore, in comparison with the model trained on MultiWOZ 2.1, our NLU model presents convincing performances such as to constitute a new benchmark for the Italian NLU. Our work demonstrate not only the issues related to the training of NLU models on low resource language, but, more importantly, constitutes a starting point for working on Italian models, specifically pre-trained on dialogic dataset like JILDA . For future work, we will look into pretraining the LMs on more dialogic data such as Italian Reddit.

## References


Siqi Bao, Huang He, Fan Wang, Hua Wu, and Haifeng Wang. 2020. PLATO: Pre-trained dialogue generation model with discrete latent variable. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 85–96, Online. Association for Computational Linguistics.

Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. 2018. MultiWOZ - a large-scale multi-domain wizard-of-Oz dataset for task-oriented dialogue modelling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5016–5026, Brussels, Belgium. Association for Computational Linguistics.

Giuseppe Castellucci, Valentina Bellomaria, Andrea Favalli, and Raniero Romagnoli. 2019. Multi-lingual intent detection and slot filling in a joint bert-based model. *In ArXiv abs/1907.02884.*

Saurabh Chakravarty, Raja Venkata Satya Phanindra Chava, and Edward Fox. 2019. Dialog acts classification for question-answer corpora. In *ASAIL@ICAIL*.

Marcus Colman, Arash Eshghi, and Pat Healey. 2008. Quantifying ellipsis in dialogue: an index of mutual understanding. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pages 96–99, Columbus, Ohio. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Mihail Eric, Rahul Goel, Shachi Paul, Abhishek Sethi, Sanchit Agarwal, Shuyang Gao, Adarsh Kumar, Anuj Goyal, Peter Ku, and Dilek Hakkani-Tur. 2020. MultiWOZ 2.1: A consolidated multi-domain dialogue dataset with state corrections and state tracking baselines. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 422–428, Marseille, France. European Language Resources Association.

Arash Eshghi, Igor Shalyminov, and Oliver Lemon. 2017. Interactional dynamics and the emergence of language games. *CEUR Workshop Proceedings*, 1863:17–21.

Raquel Fernández and Jonathan Ginzburg. 2002. Non-sentential utterances: A corpus-based study. *Traitement Automatique des Langues*, 43(2).

Ting Han, Ximing Liu, Ryuichi Takanabu, Yixin Lian, Chongxuan Huang, Dazhen Wan, Wei Peng, and Minlie Huang. 2021. Multiwoz 2.3: A multi-domain task-oriented dialogue dataset enhanced with annotation corrections and co-reference annotation. In *Natural Language Processing and Chinese Computing*, pages 206–218, Cham. Springer International Publishing.

Julian Hough. 2015. *Modelling Incremental Self-Repair Processing in Dialogue*. Ph.D. thesis, Queen Mary University of London.

Raymond Li, Samira Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018. Towards deep conversational recommendations. *In Advances in Neural Information Processing Systems 31 (NIPS 2018)*, pages 9748–9758.

Xingkun Liu, Arash Eshghi, Pawel Swietojanski, and Verena Rieser. 2021. *Benchmarking Natural Language Understanding Services for Building Conversational Agents*, pages 165–183. Springer Singapore, Singapore.

Ryan Lowe, Nissan Pow, Iulian Vlad Serban, Chia-Wei Liu, and Jielle Pineau. 2017. Training end-to-end dialogue systems with the ubuntu dialogue corpus. *Dialogue and Discourse*, 8(1):31–65.

Nadia Mana, Roldano Cattoni, Emanuele Pianta, Franca Rossi, Fabio Pianesi, and Susanne Burger. 2004. The italian nespole! corpus: a multilingual database with interlingua annotation in tourism and medical domains. *In Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04).*

Shikib Mehri, Evgeniia Razumovskaia, Tiancheng Zhao, and Maxine Eskenazi. 2019. Pretraining methods for dialog context representation learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3836–3845, Florence, Italy. Association for Computational Linguistics.

Sinarwati Mohamad Suhaili, Naomie Salim, and Mohamad Nazim Jambli. 2021. Service chatbots: A systematic review. *Expert Systems with Applications*, 184:115461.

Bill Noble and Vladislav Maraev. 2021. Large-scale text pre-training helps with dialogue act recognition, but not without fine-tuning. In *Proceedings of the 14th International Conference on Computational Semantics (IWCS)*. Association for Computational Linguistics.

Marco Polignano, Pierpaolo Basile, Marco de Gemmis, Giovanni Semeraro, and Valerio Basile. 2019. AlBERTo: Italian BERT Language Understanding Model for NLP Challenging Tasks Based on Tweets. In *Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019)*, volume 2481. CEUR.

Matthew Purver, Christine Howes, Eleni Gregoromichelaki, and Patrick G. T. Healey. 2009. Split utterances in dialogue: A corpus study. In *Proceedings of the 10th Annual SIGDIAL Meeting on Discourse and Dialogue (SIGDIAL 2009 Conference)*, pages 262–271, London, UK. Association for Computational Linguistics.

Lance Ramshaw and Mitch Marcus. 1995. Text chunking using transformation-based learning. In *Third Workshop on Very Large Corpora*.

Eugénio Ribeiro, Ricardo Ribeiro, and David Martins de Matos. 2019. Deep dialog act recognition using multiple token, segment, and context information representations. *J. Artif. Intell. Res.*, 66:861–899.

Stefan Schweter. 2020. Italian bert and electra models.

Elizabeth Shriberg. 1996. Disfluencies in switchboard. In *In Proceedings of the International Conference on Spoken Language Processing*, volume 96, pages 3–6. Citeseer.

Irene Sucameli, Alessandro Lenci, Bernardo Magnini, Maria Simi, and Manuela Speranza. 2020. Becoming jilda. *In Proceedings of the Seventh Italian Conference on Computational Linguistics CLIC-it 2020*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, undefinedukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010, Red Hook, NY, USA. Curran Associates Inc.

Qi Zhu, Zheng Zhang, Yan Fang, Xiang Li, Ryuichi Takanobu, Jinchao Li, Baolin Peng, Jianfeng Gao, Xiaoyan Zhu, and Minlie Huang. 2020. ConvLab-2: An open-source toolkit for building, evaluating, and diagnosing dialogue systems. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 142–149, Online. Association for Computational Linguistics.

# 7 Appendix

## 7.1 Hyper-parameter tuning procedure

We tried 12 different hyperparameter combinations on the validation set: three batch size values (32, 64, 128) and four learning rates ($1e-4$, $2e-5$, $3e-4$, and $5e-5$). Moreover, we kept the number of steps low to prevent overfitting, with check-step: 300 and max-step: 3000. The other relevant settings include *finetune*, *context* and *context-grad*. The fist one determines if the model will be tuned or not with the BERT parameter. If *fine-tune:false*, only added classification layers will be tuned.

The context parameter defines if use context information. If context: false, the [CLS] representation of the single utterance is passed to the intent classifier while the tokens' representations are passed to the slot classifier. If true, context utterances of the last three turns are concatenated and provide context information with embedding of [CLS] for dialogue act and slot classification.

Finally, context-grad determines whether compute the gradient through context representation, and then back-propagate the loss to the context encoder.

According to the results obtained evaluating the results on the validation set, we fixed the hyper-parameters as follows:

```
"model": {
"finetune": true,
"context": true,
"context_grad": false,
"check_step":300,
"max_step":3000,
"batch_size": 64,
"learning_rate": 1e-4,
"adam_epsilon": 1e-8,
"warmup_steps": 0,
"weight_decay": 0.0,
"dropout": 0.1,
"hidden_units": 768 }
```

## 7.2 DAs and slots occurrences in JILDA

Table below reports the number of Dialogue acts' and slots' occurrences in the JILDA dataset. As shown in the Table, some DAs and slots are higher represented than other; the higher is their representation in the dataset, the more accurate the models' classification is, as discussed in Section 5.

| | Label | Occurences |
|---|---|---|
| DA | greet | 6.140 |
| | deny | 2.016 |
| | select | 890 |
| | inform | 14.538 |
| | request | 9.434 |
| Slot | age | 130 |
| | area | 1.472 |
| | company-name | 556 |
| | company-size | 732 |
| | contact | 827 |
| | contract | 1.486 |
| | degree | 1.243 |
| | duties | 1.741 |
| | job-description | 1.362 |
| | languages | 1.085 |
| | location | 1.922 |
| | other | 559 |
| | past-experience | 882 |
| | skill | 1.994 |

Table 6: DA' and slots' occurrences in JILDA .