Research Statement – Edoardo Maria Ponti

The overarching goal of my research is to endow machines with a human-like propensity to acquire *any* language quickly from limited examples and to use language creatively and flexibly in unforeseen circumstances. The predominant paradigm of machine learning, on the contrary, is arguably data hungry and relies on the overlap between train and evaluation data distributions. To narrow this chasm, I take inspiration from the abstract mechanisms that underpin the unique linguistic abilities of humans. For instance, if we heard the word *gavagai* from the Arunta language uttered upon noticing a RABBIT, we would surmise that it refers to the whole animal, and not to a body part or the colour of its fur. Machines could learn in a **sample-efficient** fashion by virtue of an appropriate **inductive bias** which favours specific hypotheses among those compatible with the evidence. Moreover, if we mastered how to answer questions in Maricopa and tell stories in Dyirbal, then we should be able to answer questions in Dyirbal without trouble. Machines could achieve **systematic generalisation** and robustness by disentangling autonomous facets of knowledge (**modules**) and recombining them in original ways.

Overcoming the challenges of sample efficiency and systematic generalisation is equally important in view of its practical ramifications. Specifically, it would enable truly **multilingual natural language processing** applications. In fact, most of the world's languages suffer from the paucity of labelled data and display a considerable degree of variation. This curbs the effectiveness of traditional supervised learning, as it is arduous both to find in-domain data and to bridge between differently distributed domains in this scenario. My current mission, therefore, consists in innovating algorithms and creating resources that facilitate **cross-lingual transfer and multi-task learning**. Ultimately, these endeavours hold promise to mitigate the current digital divide, giving access to vital services to communities speaking endangered, under-documented, and minority languages.

Past Research and Doctoral Dissertation

Inductive Biases to Learn Quickly

An adequate inductive bias that expedites language acquisition can be constructed by adopting a Bayesian perspective towards learning. In [9, 5], I inferred (through Laplace and variational approximations) a distribution over neural parameters *and* architectures from data in multiple resource-rich languages and leveraging features from typological databases as side information. This distribution, imbued with universal linguistic knowledge, subsequently served as a **Bayesian prior** for few-shot updates in a variety of languages and tasks (such as character-level language modelling and commonsense reasoning), yielding large gains in sample efficiency and performance.

Nevertheless, a similarly constructed inductive bias is bound to merely reflect the properties of source languages, which are often not identically distributed as the target languages. This makes methods like meta-learning, which hinge upon the exchangeability of these two distributions, brittle when adapting to outlier languages. To counter this, in [6] I recast **meta-learning as a competitive game** in two flavours: 1) a minimax formulation minimises the loss of languages chosen adversarially; 2) a Neyman–Pearson formulation places constraints on optimisation whereby the loss for

a subset of languages cannot exceed a predefined threshold. This increases the robustness of the prior induced via meta-learning under distribution shifts.

Furthermore, the inductive bias does not need to originate exclusively from natural languages. Concurrent work found that modelling music and code beforehand accelerates language learning in neural networks. However, these sources of inductive bias fail to account for the grounded and inter-personal nature of language. For this reason, in [1] I proposed the idea of exploiting instead **emergent languages**, which arise when artificial agents need to communicate to solve referential games based on images. Specifically, I pre-trained models on unsupervised referential games and then fine-tuned them on natural language tasks, such as few-shot neural machine translation. Intuitively, transfer succeeds because, while the lexicon is mostly arbitrary, the semantic and syntactic properties useful for grounded communication are shared by both emergent and natural languages.

Modular Design for Systematic Generalisation

Another key aspect of the human language faculty is the **integrated but modular** system in which it is deeply entrenched. To mirror this, I devised models with a similar **design**. In [10], I argued that, in order to achieve systematic generalisation, each task should be conceived as the combination of autonomous skills (e.g., related to a specific application, language, or modality) needed to solve it, rather than a monolith where all knowledge is lumped together into a single set of neural parameters. Under this assumption, **the space of neural parameters is structured and can be factorised** into distinct latent variables for each skill. This defines a generative Bayesian model, which can be implemented in practice through hyper-networks that take representations of the appropriate skill combinations in input and yield conditional distributions of neural parameters in output. This approach demonstrated enhanced generalisation capabilities in part-of-speech tagging and named entity recognition across a wide array of typologically diverse languages.

Creation and Translation of Multilingual Resources

In synergy with efficient and robust algorithms, the creation of gold multilingual data equally sustains multilingual natural language processing. I contributed to these annotation efforts by releasing three multilingual datasets, MultiSimLex [11] for lexical semantic similarity, XCOPA [7] for commonsense causal reasoning, and AM^{2}iCo [3] for word meaning in context. These are designed to **maximise language diversity** in terms of typological features, family membership, and areal dispersion. As a result, they include extremely low-resource languages such as Haitian Creole and Southern Quechua. Moreover, their data is curated to **avoid cultural bias**, as salient concepts and typical situations vary dramatically across the world.

An alternative procedure to obtain data is to translate examples from resource-rich languages. This provides abundant supervision to subsequently train a classifier in a resource-poor language. While effective, this pipeline suffers from some limitations: 1) the errors that accumulate along its steps cannot be corrected; 2) only the maximum-likelihood translation is generated, which may fail to fully render the original meaning. In [8], I integrated both translator and classifier into a single joint model, treating the **intermediate translations as a latent random variable**. By performing inference under this model, 1) the translator can be fine-tuned end-to-end based on the classification loss; 2) multiple translation samples can be drawn to perform ensemble prediction in the downstream task. This produced state-of-the-art results for several multilingual NLU tasks.

Current and Future Agenda

Discovery of Linguistic Skills

Part of my ongoing research builds upon my work on modularity for systematic generalisation. In [10], I assumed the identity and granularity of skills, as well as their task-specific combinations, to be given. In more naturalistic settings like grounded language learning, however, we do not know these details in advance. Thus, I aim to **learn linguistic skills**, as well as their combinations upon which neural parameters depend, **end-to-end**. I obtained promising early results by evaluating this approach on BabyAI, a platform where agents have to follow linguistic commands in a simulated environment. An example is offered in the figure above: actions are chosen not solely according to the visual and linguistic inputs, but also conditioned on the set of inferred task-specific skills.

In addition, latent skills can also be discovered **in an unsupervised fashion**. Given the gargantuan scale of pre-trained language models, a recent trend is keeping their parameters frozen while only fine-tuning prompts, special embeddings prepended to the encoder input (and decoder output). This setting resembles a Markov Decision Process, where a state (input text) and an action (prompts) result in a new state (output text) according to a transition probability (pre-trained language model). Drawing from '**empowerment**' in reinforcement learning, I intend to maximise the mutual information between latent skills, upon which prompts (actions) are conditioned, and input-output pairs (state trajectories). This will favour the discovery of skills that 1) are diverse and 2) have high control on the outputs.

Multilingual Vision and Dialogue

More generally, in line with my long-standing goal of better aligning language learning in machines with humans, another aim for my future research is to explore multilinguality beyond text. In fact, models should better reflect how linguistic communication is 'situated' in a real environment (vision) and embodied in an agent capable of interaction (dialogue). During my postdoc, I am establishing a footing in both areas. In [2], I performed **visually grounded multilingual reasoning**, revealing how Anglo-centric training data leads to almost random performance on evaluation data created to faithfully represent other languages and cultures.

In a recent survey on **multilingual task-oriented dialogue** [4], I argued that multilinguality poses unique challenges for dialogue systems, such as few-shot learning of end-to-end models, fluency in (possibly morphologically rich) generated text, and subject recruitment for human-centred evaluation. This laid the groundwork for expanding language coverage and dialogue capabilities of current task-oriented dialogue systems in future research. I have been jointly awarded a European Research Council Proof-of-Concept grant as a means to achieve this ambitious goal.

I am looking forward to exploring these research directions together with my prospective group of PhD students.

References

[1] Yaoyiran Li, **Edoardo Maria Ponti**, Ivan Vulić, and Anna Korhonen. 2020. Emergent communication pretraining for few-shot machine translation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4716–4731.

[2] Fangyu Liu, Emanuele Bugliarello, **Edoardo Maria Ponti**, Siva Reddy, Nigel Collier, and Desmond Elliott. 2021. Visually grounded reasoning across languages and cultures. In *2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online. Under submission.

[3] Qianchu Liu, **Edoardo Maria Ponti**, Diana McCarthy, Ivan Vulić, and Anna Korhonen. 2021. AM2iCo: Evaluating word meaning in context across low-resource languages with adversarial examples. In *2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online. Under submission.

[4] Evgeniia Razumovskaia, Goran Glavaš, Olga Majewska, **Edoardo Maria Ponti**, Anna Korhonen, and Ivan Vulić. 2021. Crossing the conversational chasm: A primer on multilingual task-oriented dialogue systems. *arXiv preprint arXiv:2104.08570*.

[5] **Edoardo Maria Ponti**. 2021. *Inductive Bias and Modular Design for Sample-Efficient Neural Language Learning*. Ph.D. thesis, University of Cambridge.

[6] **Edoardo Maria Ponti**, Rahul Aralikatte, Disha Shrivastava, Siva Reddy, and Anders Søgaard. 2021. Minimax and Neyman–Pearson meta-learning for outlier languages. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1245–1260, Online.

[7] **Edoardo Maria Ponti**, Goran Glavaš, Olga Majewska, Qianchu Liu, Ivan Vulić, and Anna Korhonen. 2020. XCOPA: A multilingual dataset for causal commonsense reasoning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2362–2376, Online.

[8] **Edoardo Maria Ponti**, Julia Kreutzer, Ivan Vulić, and Siva Reddy. 2021. Modelling latent translations for cross-lingual transfer. *arXiv preprint arXiv:2107.11353*.

[9] **Edoardo Maria Ponti**, Ivan Vulić, Ryan Cotterell, Roi Reichart, and Anna Korhonen. 2019. Towards zero-shot language modeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2900–2910, Hong Kong, China.

[10] **Edoardo Maria Ponti**, Ivan Vulić, Ryan Cotterell, Marinela Parovic, Roi Reichart, and Anna Korhonen. 2021. Parameter Space Factorization for Zero-Shot Learning across Tasks and Languages. *Transactions of the Association for Computational Linguistics*, 9:410–428.

[11] Ivan Vulić, Simon Baker, **Edoardo Maria Ponti**, Ulla Petti, Ira Leviant, Kelly Wing, Olga Majewska, Eden Bar, Matt Malone, Thierry Poibeau, Roi Reichart, and Anna Korhonen. 2020. Multi-SimLex: A large-scale evaluation of multilingual and crosslingual lexical semantic similarity. *Computational Linguistics*, 46(4):847–897.