Soft Skills in the Wild: Challenges in Multilingual Classification

Laura Vásquez-Rodríguez[1], Bertrand Audrin[2], Samuel Michel[1], Samuele Galli[3], Julneth Rogenhofer[2], Jacopo Negro Cusa[3], Lonneke van der Plas[4,1]

[1]Idiap Research Institute, Switzerland
[2]EHL Hospitality Business School, HES-SO,
University of Applied Sciences and Arts Western Switzerland, Switzerland
[3]Arca24.com SA, Switzerland
[4]Università della Svizzera italiana, Switzerland

Correspondence: laura.vasquez@idiap.ch, bertrand.audrin@ehl.ch, lonneke.vanderplas@usi.ch

Abstract

Soft skills are a crucial factor in candidate selection for recruitment. However, they are often overlooked due to the challenges in their identification. In this study, we compare soft and hard skills as well as occupations, both in terms of surface and semantic properties of the annotations and as part of an automatic extraction task, showing clear differences between the types of skills. Soft skills can be easily limited to a small number of categories, as we show in our annotation framework, which is based on well-known taxonomies. However, the way they are expressed in texts varies more widely than other entity types. These insights help to understand possible causes for the large variation in performance we see when using a multilingual BERT-based classifier for the identification of soft skills compared to other entities, which can help the community to develop more reliable algorithms for recruitment.[1]

1 Introduction

Applicant Tracking Systems (ATS) have often focused on hard skills only and neglected soft skills in their matchmaking of candidates with job openings. The notion of soft skills refers to behavioral and social abilities that people tend to possess or develop through social interactions (Heckman and Kautz, 2012). There is a debate about what these skills entail and how to label them, which, of course, makes their identification very complex, as the way they are expressed may vary from person to person.

Research on skill extraction has predominantly focused on identifying occupations or hard skills (Senger et al., 2024). Occupations are very straightforward and refer to a large set of clearly identifiable positions (e.g., "plumber" or "architect"). Hard skills are also quite straightforward to ground in existing knowledge and categorized according to employment, but many different hard skills can be relevant for each specific position (e.g., a plumber needs to have a specific skill set: blueprint comprehension, pipe installation, drilling, etc.). In that respect, there are many more hard skills than occupations. Soft skills, for their part, also known as behavioral skills (Tamburri et al., 2020), are primarily acquired in social contexts and may be independent of technical knowledge (Sayfullina et al., 2018). There is a limited set of soft skills that exist, and these skills can be relevant for a variety of jobs (e.g., "collaboration" can be useful for a plumber or an architect). Despite their importance, their identification and impact remain challenging to assess due to their abstract nature, which can explain why less attention has been given to soft skills. Specific findings remain scarce and domain-specific, often differing significantly in data, methodology, and language (Sayfullina et al., 2018; Beauchemin et al., 2022; Zhang et al., 2022).

One of the main challenges is that these approaches have mostly focused on extracting soft skills from job advertisements (also referred to as job offers). However, soft skills in job advertisements are more likely to be standardized and are an opportunity for the company to develop its employer branding (Elving et al., 2013). In contrast, soft skills may appear very differently in resumes, sometimes less explicitly or straightforwardly, if at all, especially for certain roles for which it is not common for candidates to emphasize this part of their skill set. Moreover, soft skills are often identified in later stages of recruitment, typically through interviews or work simulations.

Our study is developed within the context of SEM24 project, which is supported by the Innosuisse (Swiss Innovation Agency). This project aims to enhance multilingual, multidomain competency detection in the European job market, with a particular focus on developing explainable algorithms for fairer recruitment processes. In this study, we

[1] We will release our results on GitHub: `https://github.com/idiap/multilingual_skill_extraction`

Surface Properties →		Total Entities			Unique Entities			Unique/Total Ratio			Avg. Len		
Text ↓	Language	Hard	Occ	Soft	Hard	Occ	Soft	Hard	Occ	Soft	Hard	Occ	Soft
Jobs (our annotations)	EN	878	322	326	777	109	255	0.88	0.34	0.78	37.53	22.19	30.36
	FR	1119	126	777	886	107	282	0.79	0.85	0.36	30.67	29.13	19.61
	IT	704	116	285	651	101	216	0.92	0.87	0.76	53.67	22.52	34.30
	PT	667	81	257	560	68	89	0.84	0.84	0.35	30.87	21.37	15.42
Jobs (Sayfullina)	EN	-	-	7403	-	-	1140	-	-	0.15	-	-	13.87
Jobs (Fijo)	FR	-	-	932	-	-	702	-	-	0.75	-	-	58.67
Jobs (Green)	EN	12573	2571	-	10079	1591	-	0.80	0.62	-	32.67	17.49	-
Resumes (our annotations)	EN	4024	692	520	3020	565	351	0.75	0.82	0.68	20.01	24.85	16.52
	FR	2063	645	441	1700	466	294	0.82	0.72	0.67	25.76	21.66	18.63
	IT	1985	645	464	1613	435	318	0.81	0.67	0.69	26.77	21.91	22.92
	PT	3312	729	200	2439	447	126	0.74	0.61	0.63	28.79	24.78	19.60

Table 1: We report the number of total and unique entities and its ratio, and average character length of hard skills, occupations, and soft skills in resumes and job offers.

specifically examine the detection of soft skills in multilingual environments. The key contributions of this paper are:

1. A discussion of the varying nature of soft skills across different document types and languages.
2. A comparative evaluation of multilingual soft skills identification from job offers and resumes.
3. Development of an annotation framework for identifying soft skills in multilingual job offers and resumes. To the best of our knowledge, this is the first time the O*NET resource has been leveraged for guidelines design in the soft skill extraction task, extending its use beyond a taxonomy of occupational terms.

2 Methodology

The main objective of this research is to characterize and understand the variable nature of soft skill annotations in job offers and resumes in a multilingual setting. As a first step, we collected multilingual resumes and job offers (Section 2.1) and annotated relevant entities following in-house annotation guidelines (Section 2.2). We then analyzed the extracted entities, characterizing the datasets and conducting a semantic analysis to explore the main differences between job offers and resumes across different languages (Section 2.3). Finally, we assessed the impact of annotation variability on performance through experiments on an entity-based classification task (Section 2.4).

2.1 Data

We collected a total of 800 resumes and job offers in 4 different languages: English (EN), French (FR), Italian (IT), and Portuguese (PT), and multiple domains such as engineering, administration, and management from our industrial partner of the SEM24 project. All documents were annotated by HR specialists according to the annotation guidelines (See Section 2.2) using a span-based Named Entity Recognition (NER) approach, where relevant entities are sequences of multiple tokens that are explicitly mentioned in the text. The annotation task was performed using the Docanno (Nakayama et al., 2018) tool. Native speakers annotated texts in French and Italian, whereas annotators with C2 proficiency annotated the other languages.

2.2 Annotations Guidelines

We extended the annotation guidelines proposed by Vásquez-Rodríguez et al. (2024) to include soft skills. Two HR researchers elaborated the proposed guide based on the O*NET taxonomy (Peterson et al., 2001) and the HEXACO personality inventory (Ashton and Lee, 2007). The categorization and identification of soft skills followed three predefined categories, then divided into a total of 21 subcategories as follows:[2]

- **Social Skills:** Coordination, Instructing, Negotiation, Persuasion, Service Orientation, and Social Perceptiveness.
- **Thinking Skills:** Active Learning, Active Listening, Complex Problem Solving, Critical Thinking, Judgment and Decision Making, Learning Strategies, Monitoring, and Time Management.

[2] These categorizations were used as a guide for the annotators to define a clear criterion that could define more precisely the concept of soft skills. However, the final labeling of entities for training was unified into the "Soft Skill" label.

Evaluation →			Exact (F1-score)			Partial (F1-score)		
Skills Type	Dataset ↓	Language	Precision	Recall	F1	Precision	Recall	F1
Jobs (Our annotations)	Soft Skills	EN	0.320	0.348	0.333	0.440	0.478	0.458
		FR	0.774	0.818	**0.796**	0.823	0.869	**0.845**
		IT	0.393	0.393	0.393	0.625	0.625	0.625
		PT	0.750	0.581	0.655	0.812	0.629	0.709
Jobs (Sayfullina)		EN	0.879	0.887	**0.883**	0.921	0.93	**0.926**
Jobs (Fijo)		FR	0.354	0.429	0.388	0.54	0.655	0.592
Jobs (Our annotations)	Hard Skills	EN	0.286	0.306	0.294	0.460	0.507	0.480
		FR	0.399	0.465	**0.427**	0.551	0.641	**0.589**
		IT	0.241	0.263	0.251	0.478	0.523	0.499
		PT	0.395	0.495	**0.438**	0.512	0.644	**0.569**
Resumes (Our annotations)	Soft Skills	EN	0.415	0.347	0.378	0.537	0.449	0.489
		FR	0.575	0.455	**0.508**	0.675	0.535	0.597
		IT	0.525	0.544	**0.534**	0.636	0.658	**0.647**
		PT	0.158	0.200	0.176	0.263	0.333	0.294
	Hard Skills	EN	0.354	0.356	0.355	0.490	0.491	0.490
		FR	0.352	0.389	0.369	0.501	0.555	0.526
		IT	0.411	0.426	0.418	0.577	0.600	0.588
		PT	0.449	0.562	0.497	0.548	0.688	**0.608**

Table 2: We report the exact (i.e., the entire entity was detected) and partial (i.e., entity was detected partially) scores for the soft and hard skills detection of the multilingual BERT model.

- **Personality Traits:** Achievement Orientation, Adjustment, Conscientiousness, Independence, Interpersonal Orientation, Practical Intelligence, and Social Influence.

Before the annotation process, all annotators were trained during an in-person workshop to discuss the final annotation guidelines and solve any disagreements between the participants. The strengths of our annotation approach lie in the fact that the annotation guidelines were developed by HR researchers following clear guidelines based on reliable frameworks. We further differ from previous work because we do not follow a static taxonomy of concepts that are expected to be explicit in the text (Sayfullina et al., 2018), and annotators are not limited to any particular domain (Beauchemin et al., 2022).

2.3 Surface Properties and Semantic Analysis

As for the surface properties of texts, we report the number of tokens and unique types (including the ratio between these metrics) of hard skills, soft skills, and occupations in the annotations in Table 1. Also, we calculated the average length of the soft skills measured by the number of characters. For all the metrics, we report the results by document type (i.e., resumes vs job offers) and language.

Another relevant aspect is the semantic similarity between soft skills in different contexts. For this analysis, we compare the extracted soft skills using the t-SNE algorithm (van der Maaten and Hinton, 2008). This algorithm reduces high-dimensional embeddings into a lower-dimensional space, where similar data points are grouped based on local similarities in their original space. The visualization of clusters shows skills that are potentially equivalent within the selected samples of documents. We highlight our results based on three different scenarios: job offers vs resumes, by language, and both resumes/job offers by language. We encoded the extracted soft skills using multilingual[3] SBERT pretrained embeddings (Reimers and Gurevych, 2020). These embeddings were then input into the t-SNE algorithm[4] using the scikit-learn Python library (Pedregosa et al., 2011). To visualize potential clusters in the data, we experimented with various perplexity levels (i.e., 10, 30, 50, 70, 100), with 50 yielding the best results across all our experiments.

2.4 Skill extraction experiments

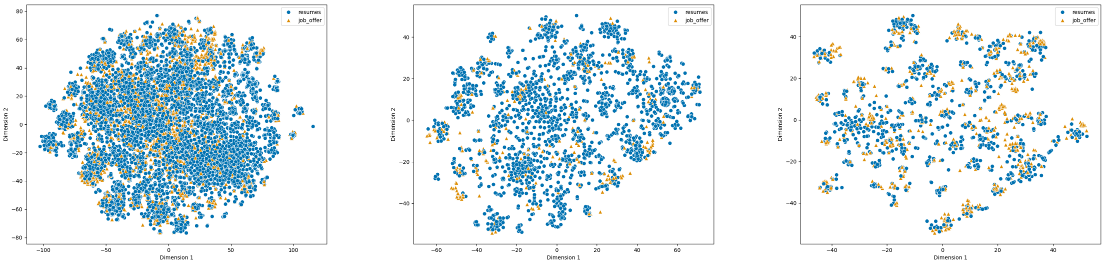
To explore the impact of soft skills annotation variability on performance (measured by F1-score) in the skill extraction task, we trained a supervised system for token classification using a span-based approach.[5] We employed a BERT-based multilingual model[6] and fine-tuned it using our manually annotated skill datasets. The corpus was split into

[3] https://huggingface.co/sentence-transformers/distiluse-base-multilingual-cased-v1

[4] https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html

[5] We acknowledge that skills can sometimes be inferred from text, and the span-based approach may not always be the most suitable. However, we chose to leverage existing online resources and tools mostly designed for a span-based approach.

[6] https://huggingface.co/google-bert/bert-base-multilingual-uncased



(a) Hard Skills

(b) Occupations

(c) Soft Skills

Figure 1: t-SNE visualization of all entities embeddings, comparing job offers and resumes.

train, validation, and test subsets (80/10/10) within each language and document type. In addition, we trained the model on datasets available from previous work, Sayfullina (Sayfullina et al., 2018) for soft skills in English, Fijo (Beauchemin et al., 2022) for French,[7] and the Green dataset (Green et al., 2022) for hard skills in English. We differ from previous work in soft skills, so that we rely on the knowledge of the HR researchers and detailed guidelines for the identification of entities in the text, rather than following a limited taxonomy of concepts (Sayfullina et al., 2018). Similarly, our proposed categories are based on updated resources that are more specific and still relevant, and available today. For evaluation, we post-processed the results using the IOB format (Ramshaw and Marcus, 1999) and then calculated precision, recall, and F1-score using the *nervaluate* Python library.[8]

3 Results

We present the surface-level statistics of soft skills, hard skills, and occupations across all datasets in Table 1. In addition, Figure 1 shows a semantic analysis comparing soft skills to hard skills and occupations. This visualization highlights how some concepts, such as hard skills, tend to form many clusters more densely scattered over the space compared to soft skills, which show a more clearly delineated representation of fewer clusters.

Furthermore, we include the main results of our evaluation in Table 2. These results represent a comparative evaluation of the skill extraction task for soft skills for both existing and in-house datasets. Also, we add the extraction results for hard skills as a reference, not only across document types but also across languages. Finally, in Table 3, we include examples of system outputs (i.e., soft skills) for a closer look at the variability between the datasets and their impact on prediction.

4 Discussion

The quality of automatic extractions of soft skills has high variability. In Table 2, we observe how performance in soft skills (measured by exact F1-scores) varies in ranges of 0.1-0.8, while in hard skills, values are more stable (between 0.2-0.4).

For the detection of soft skills in job advertisements, it is often hard to clearly distinguish between those soft skills that pertain to the specific job at hand and others that are more used for employer branding purposes (e.g., "working in a diverse team" might refer to open-mindedness as a requirement for candidates, or as an employer branding signal that the team is diverse). In resumes, soft skills are often less standardized, and applicants are likely to use a broader variety of terms to refer to them. Moreover, soft skills are often less clearly defined, and some statements might refer to several soft skills at the same time (e.g., "ability to adapt to customers' needs" could refer to flexibility, but also to customer-centricity). This aspect of subjectivity in soft skills expression makes it a challenging task.

We observe the same variability in Table 1. There are fewer underlying categories of soft skills than occupations or hard skills, but the way they are expressed varies widely. We see more entities (both in total and unique) for soft skills in several languages and document types than the underlying 21 subcategories. The ratio between unique and total number of annotated skills shows that although soft skills are generally expressed with shorter phrases, the variability is relatively high. In Figure 1c, we see that hard skills are scattered

[7] Details of our training procedure are provided in Appendix A.1.

[8] https://pypi.org/project/nervaluate/

Dataset	Example
Jobs (Sayfullina)	R: you will have excellent communication and leadership abilities P: you will have excellent communication and leadership abilities R: be a national provider of independent sector complex healthcare P: be a national provider of independent sector complex healthcare
Jobs (Fijo)	R: anglais intermediaire connaissances excel intermediaire - avance attitude positive et aimant travailler en equipe [9] P: anglais intermediaire connaissances excel intermediaire avance attitude positive et aimant travailler en equipe R: service a la clientele traiter, analyser et gerer les correspondances aupres de la clientele interne[10] P: service a la clientele traiter, analyser et gerer les correspondances aupres de la clientele interne
Jobs (Our Annotations)	R: customer service skills great attention to detail P: customer service skills great attention to detail R: follow verbal and written instructions be able to work quickly and concisely under pressure P: follow verbal and written instructions be able to work quickly and concisely under pressure
Resumes (Our Annotations)	R: curious and thoughtful person with good inventive P: curious and thoughtful person with good inventive R: analysis of feasibility problems and ability to problem solving design P: analysis of feasibility problems and ability to problem solving design

Table 3: Soft skill detection examples for the multilingual BERT model in all datasets. For each example, we include the model's prediction (P) and the annotators' reference labels (R).

over the entire space, whereas occupations and soft skills are more clearly grouped in a smaller number of clusters. For soft skills, the clusters are more distinctly separated, supporting the idea of having fewer categories.

We cannot draw firm conclusions from the difference in performance between the languages, because here datasets also differ a lot, both in how annotations are done and what type of data is selected. For example, there is a significant gap between Sayfullina (0.883) and Fijo (0.33), not only because of the size of the dataset but also because of how soft skills are defined (e.g., "team working", "independent" vs "Être orienté vers l'action").[11] Overall it is difficult to pose strict boundaries to soft skills as in a span-based approach, which results in variable average lengths as demonstrated in Table 1. The verbosity of the Fijo datasets is also evident in Table 3, where large portions of the sentence are highlighted in both the predictions and the references. Whether a concept like "curious" or a phrase such as "ability to problem solving" is identified as a soft skill depends on multiple factors including the annotator's previous knowledge, model's learned patterns, the taxonomy design, and less evident influences such as language, writing style, document type, and context. The same variability can be found in our datasets because it is based on real-world data, where clients in one language are typically less varied than in another language.

We chose the span-based approach as it is convenient to compare against existing literature and to leverage existing annotation tools. However, conversational LLMs could help to categorize into broader, more general categories (e.g., thinking skills, social skills, personality, etc.). It could also mitigate the difficulties of extracting non-explicit skills, while making sure they align with human judgements, so that algorithms are trustworthy and reliable. We leave this avenue for future work.

5 Conclusions

In this paper, we have shown an analysis of the nature of soft skills and have provided experiments that test the performance of automatic soft skill identification, as compared to the extraction of hard skills and occupations. We demonstrated differences across skill types and across resumes as well as job offers in a number of languages, while resumes are absent in most previous work.

Although soft skills can be summarized in a small number of well-known categories, the variability in human expression is more pronounced than for hard skills. We show that this variability poses additional challenges for the extraction of soft skills when compared to hard skills. These results underline the importance of considering approaches that move away from a span-based approach. Similarly, resumes often present hard skills in lists or conjoined ways that significantly limit the ability to extract them precisely. In that respect, working on soft skills has been incredibly helpful in revealing challenges that are also faced in hard skill extraction but tend to be dismissed for simplification purposes.

[9] In English, "customer service process, analyze and manage correspondence with internal customers."

[10] In English, "intermediate english, intermediate to advanced excel skills, positive attitude and enjoys working in a team."

[11] In English, "being action-oriented."

Limitations

We recognize the importance of releasing both models and data to support the research community. However, due to privacy concerns and the intellectual property policies of the company, we are unable to release proprietary job offers and resumes. To support the reproducibility of our work, we instead provide our models and datasets based on publicly available data.

Acknowledgments

We would like to thank Arca24 HR specialists for their great effort in completing all annotations for our paper. Finally, we gratefully acknowledge the support from Innosuisse (grant 104.069 IP-ICT).

References

Michael C. Ashton and Kibeom Lee. 2007. Empirical, theoretical, and practical advantages of the hexaco model of personality structure. *Personality and Social Psychology Review*, 11(2):150–166.

David Beauchemin, Julien Laumonier, Yvan Ster, and Marouane Yassine. 2022. "fijo": a french insurance soft skill detection dataset. *arXiv*.

Wim J L Elving, Jorinde J C Westhoff, Kelta Meeusen, and Jan-Willem Schoonderbeek. 2013. The war for talent? The relevance of employer branding in job advertisements for becoming an employer of choice. *Journal of Brand Management*, 20(5):355–373.

Thomas Green, Diana Maynard, and Chenghua Lin. 2022. Development of a benchmark corpus to support entity recognition in job descriptions. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1201–1208, Marseille, France. European Language Resources Association.

James J. Heckman and Tim Kautz. 2012. Hard evidence on soft skills. *Labour Economics*, 19(4):451–464. European Association of Labour Economists 23rd annual conference, Paphos, Cyprus, 22-24th September 2011.

Hiroki Nakayama, Takahiro Kubo, Junya Kamura, Yasufumi Taniguchi, and Xu Liang. 2018. doccano: Text annotation tool for human. Software available from https://github.com/doccano/doccano.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830.

Norman G Peterson, Michael D Mumford, Walter C Borman, P Richard Jeanneret, Edwin A Fleishman, Kerry Y Levin, Michael A Campion, Melinda S Mayfield, Frederick P Morgeson, Kenneth Pearlman, et al. 2001. Understanding work using the occupational information network (o* net): Implications for practice and research. *Personnel psychology*, 54(2):451–492.

Lance A Ramshaw and Mitchell P Marcus. 1999. Text chunking using transformation-based learning. In *Natural language processing using very large corpora*, pages 157–176. Springer.

Nils Reimers and Iryna Gurevych. 2020. Making monolingual sentence embeddings multilingual using knowledge distillation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4512–4525, Online. Association for Computational Linguistics.

Luiza Sayfullina, Eric Malmi, and Juho Kannala. 2018. Learning representations for soft skill matching. In *Analysis of Images, Social Networks and Texts: 7th International Conference, AIST 2018, Moscow, Russia, July 5–7, 2018, Revised Selected Papers 7*, pages 141–152. Springer.

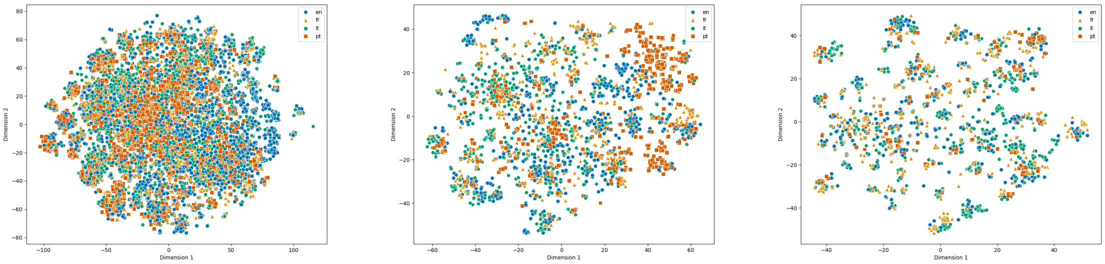
Elena Senger, Mike Zhang, Rob van der Goot, and Barbara Plank. 2024. Deep learning-based computational job market analysis: A survey on skill extraction and classification from job postings. In *Proceedings of the First Workshop on Natural Language Processing for Human Resources (NLP4HR 2024)*, pages 1–15, St. Julian's, Malta. Association for Computational Linguistics.

Damian A. Tamburri, Willem-Jan Van Den Heuvel, and Martin Garriga. 2020. Dataops for societal intelligence: a data pipeline for labor market skills extraction and matching. In *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, pages 391–394.

Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605.

Laura Vásquez-Rodríguez, Bertrand Audrin, Samuel Michel, Samuele Galli, Julneth Rogenhofer, Jacopo Negro Cusa, and Lonneke Van Der Plas. 2024. Hardware-effective approaches for skill extraction in job offers and resumes. In *RecSys in HR'24: The 4th Workshop on Recommender Systems for Human Resources, in conjunction with the 18th ACM Conference on Recommender Systems.*, pages 1–12. CEUR Workshop Proceedings.

Mike Zhang, Kristian Jensen, Sif Sonniks, and Barbara Plank. 2022. SkillSpan: Hard and soft skill extraction from English job postings. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4962–4984, Seattle, United States. Association for Computational Linguistics.

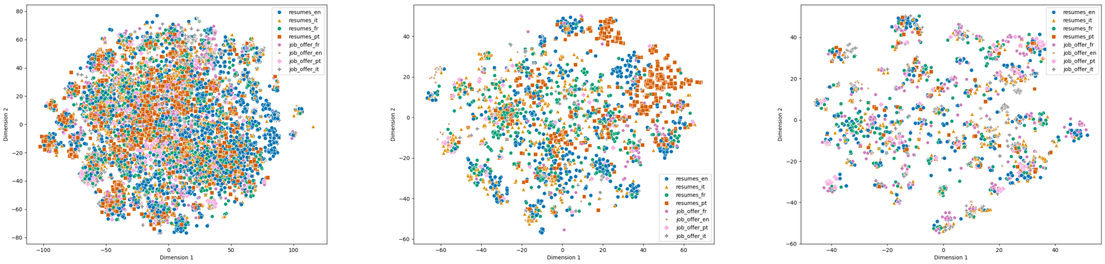


(a) Hard Skills

(b) Occupations

(c) Soft Skills

Figure 2: t-SNE visualization of all entities embeddings, comparing all languages.



(a) Hard Skills

(b) Occupations

(c) Soft Skills

Figure 3: t-SNE visualization of all entities embeddings, comparing all languages per resume and job offer.

A Appendix

A.1 Training details

As for the training parameters, we ran all experiments using 3 different seeds, then we reported the average results across all runs. The selected hyperparameters include a batch size of 16, a learning rate of 5.00×10^{-5}, and a maximum of 10 epochs. All experiments were conducted on a single NVIDIA GeForce RTX 3090 GPU with 24 GB of RAM.

A.2 Results

Further, in our analysis presented in Figure 1, we include the t-SNE visualization highlighting each language only (Figure 2) and also, considering the existing pairs of resumes-job offers and languages (Figure 3).