|  |  |  |
| --- | --- | --- |
| **Addendum**  This document provides additional information relating to both specific aspects of the post/faculty and any post specific person specification criteria. The information contained within this document should always be read in conjunction with the accompanying generic Role Profile. | | |
| **Job Title:** | Research Fellow A | |
| **Background Information/Relationships**  We are seeking qualified applicants for a position as Language Data Scientist on the ERC Synergy grant ‘NILOMORPH: The evolution of suprasegmental morphology in West Nilotic’, led by Matthew Baerman. The successful candidate will perform a key role in managing, processing and analysing language data generated across the multiple teams that make up the project. The position is based at the Surrey Morphology Group at the University of Surrey, and provides the opportunity to work in the vibrant and highly collegial research environment for which the SMG is renowned. | | |
| **Person Specification**  This section describes the sum total of knowledge, experience & competence required by the post holder that is necessary for standard acceptable performance in carrying out this role. This is in addition to the criteria contained within the accompanying generic Role Profile. | | |
|  | | **Essential/ Desirable** |
| Track record of published and/or doctoral research in their field of specialism | | **Essential** |
| Track record of both independent and team-based research. | | **Essential** |
| Competence in:   * Using continuous development and pipelines for validation, testing and publication (incl. static site generators) * Git * Programming (python, R) | | **Essential** |
| Experience with linguistic data management and creation, including but not limited to corpora and lexical resources | | **Essential** |
| Experience with large and noisy data management, data cleaning and wrangling | | **Essential** |
| Experience with Natural Language Processing for linguistics, and creation of NLP pipelines including tokenization, tagging, parsing | | **Desirable** |
| Experience with data visualisation | | **Desirable** |
| Skills in statistical analysis | | **Desirable** |
| Formal training in linguistics to degree level | | **Essential** |
| A PhD in linguistics, data science, computer science or related field | | **Essential** |
| Excellent written and spoken English | | **Essential** |
| **Key Responsibilities** | | |
| 1. Contribute to the formulation and development of tools to facilitate the analysis of large-scale cross-linguistic datasets 2. Manage the validation, versioning and archiving processes for data of different types and qualities 3. Contribute to the design and development of linguistically sophisticated synthetic datasets, and take a lead role in their delivery. 4. Analyse and scientifically interpret research results; present outcomes at scientific conferences; and prepare co-authored research publications. 5. Collaborate with other team members in an ambitious and multidisciplinary project, and participate in the vibrant, collegial activities of SMG.   **N.B. The above list is not exhaustive**. | | |