FIN-CLARIAH
2024/25

WP 4.1.6 ENRICH SURVEY DATA WITH REGISTER DATA & UNSTRUCTURED TEXT

Authors:
Adeline Clarke
University of Helsinki, Centre for Social Data Science
Maria Valaste
University of Helsinki, Centre for Social Data Science

BACKGROUND

Open-ended questions are a useful, but challenging, source of survey response data. These questions allow respondents to provide diverse answers based on their experience, and often include new answers not previously considered by researchers.

However, since these responses consist of text data and have minimal constraints on their content, open-ended questions can be difficult to analyse. There are few free tools available for survey text analysis in small languages such as Finnish. This means that open-ended questions are often underutilised in research despite their advantages, especially when the survey has many responses, making closed reading infeasible.
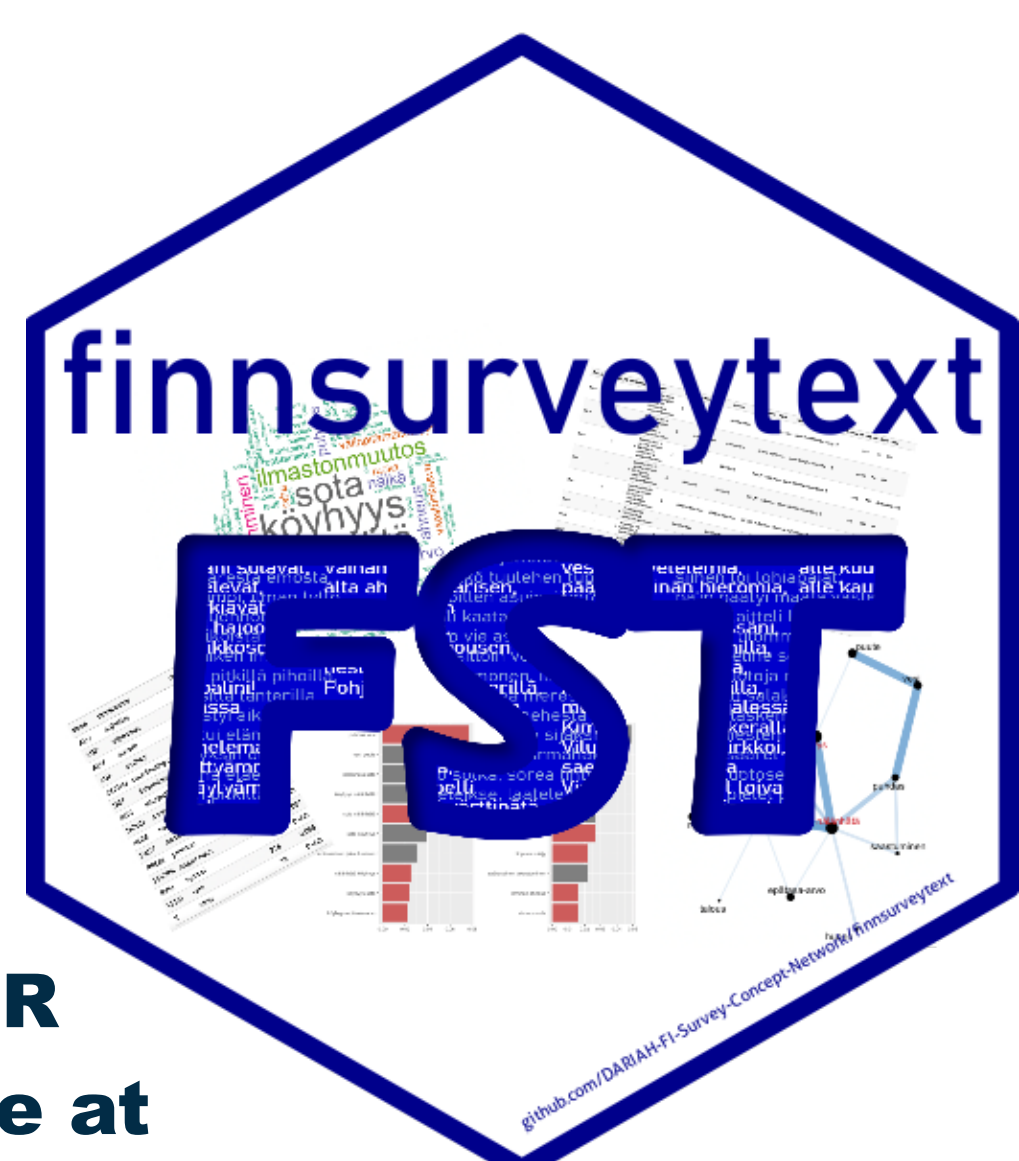
Through our package, *finnsurveytext*, we aim to lower the threshold for using open-ended survey questions in research by providing a user-friendly way for researchers to be able to begin analysing these open-ended questions without requiring costly tools or advanced programming skills.

FINNSURVEYTEXT R PACKAGE V1.0

The first version of *finnsurveytext* was released to the CRAN, R's central software repository, in February 2024. This was the result of collaboration between the University of Eastern Finland, the University of Helsinki, the University of Jyväskylä, and the University of Tampere.

The package enables users to:

- prepare and annotate text data into a standardised format ready for analysis,
- explore the data to find trends and topics,
- compare responses between different groups of participants, and,
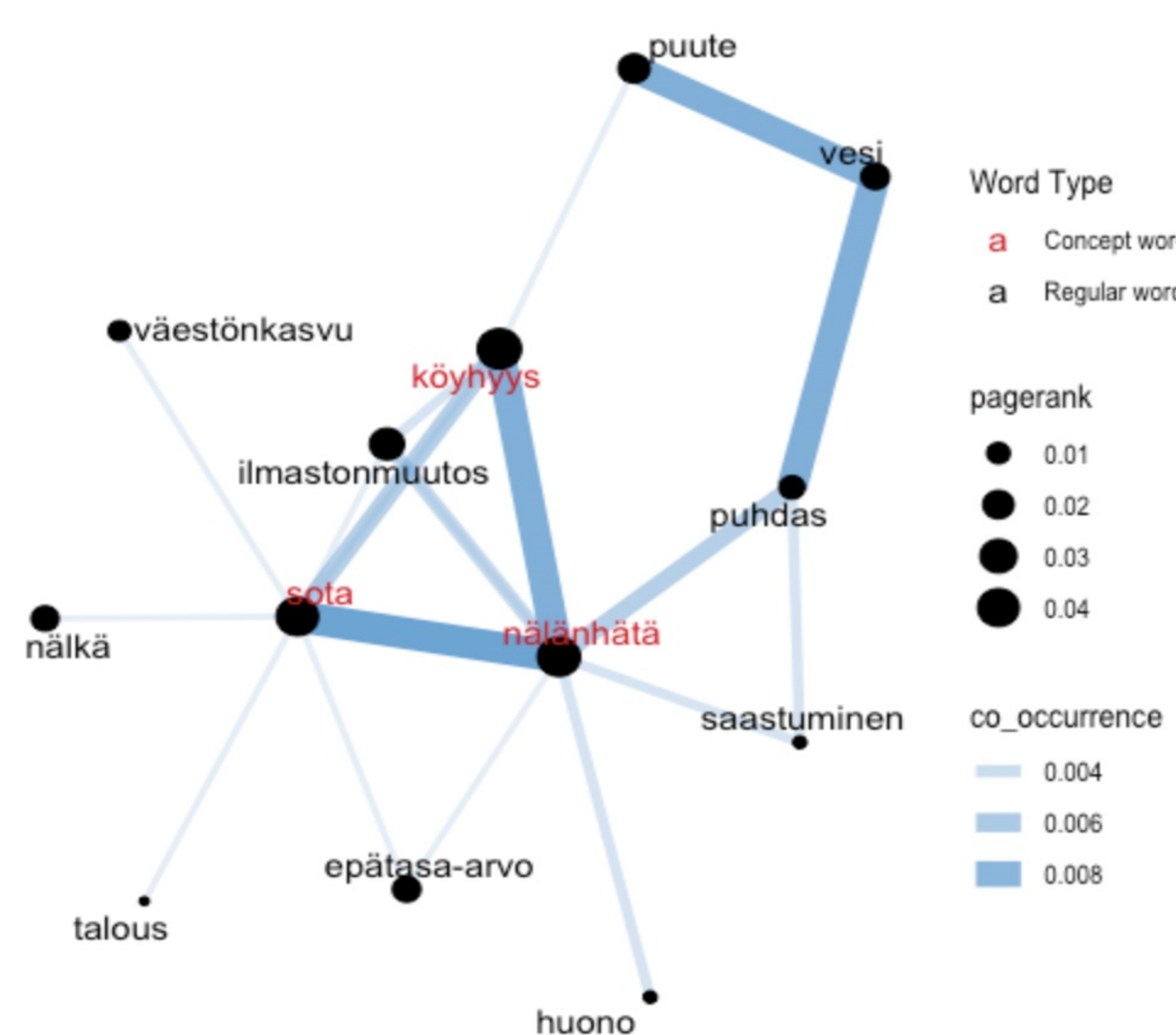- visualise these findings.

The *finnsurveytext* R package is available at
https://cran.r-project.org/package=finnsurveytext

Wordcloud created using *finnsurveytext*

The *finnsurveytext* R package can be used to produce a number of summary tables and visualisations including:

- wordclouds,
- most frequent word and n-grams tables,
- summaries of response length and word types used,
- concept network visualisations, and,
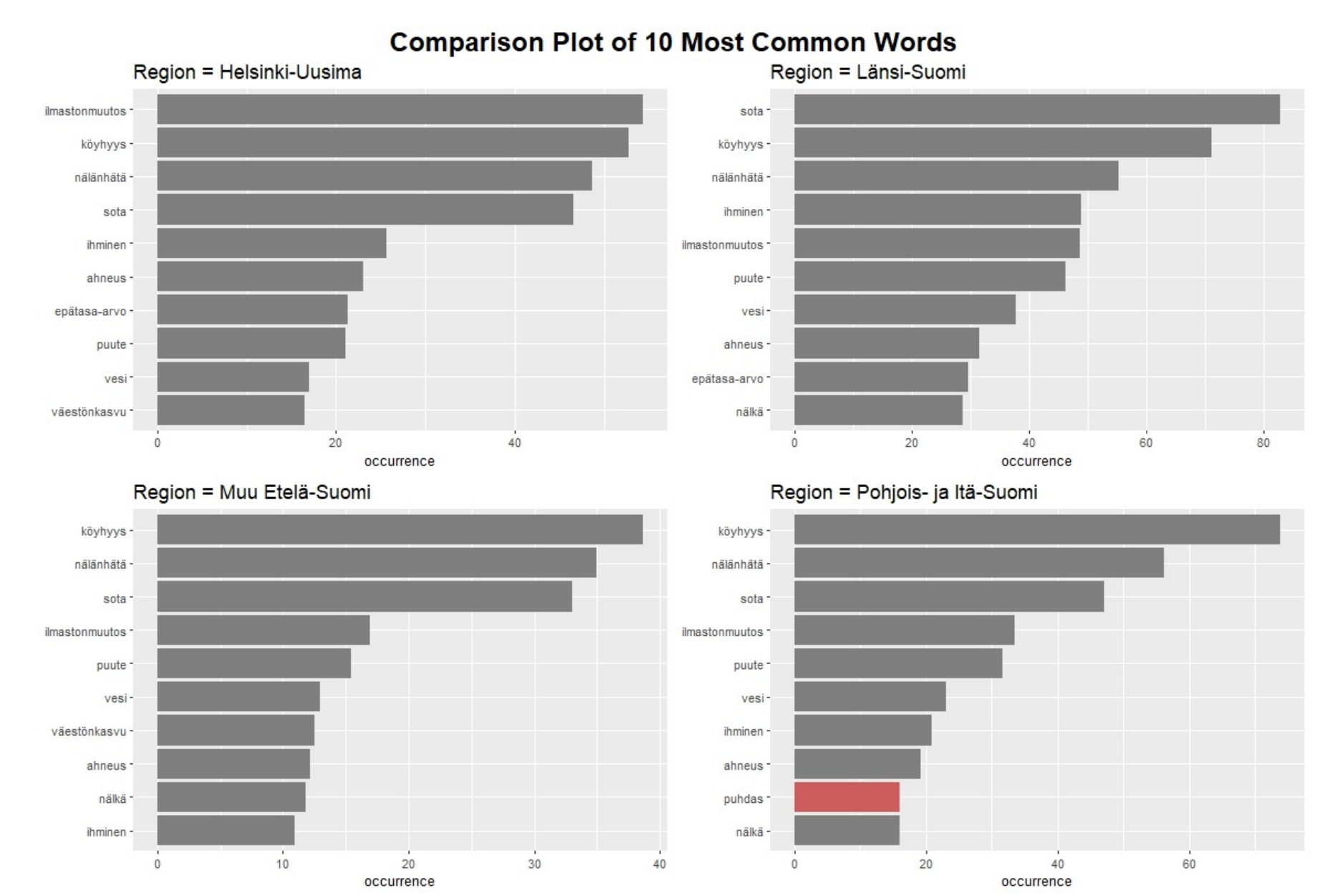- comparisons between different groups of participants.

Concept Network of words used with *köyhyys*, *nälänhätä*, and *sota*.

WP 4.1.6 PACKAGE UPDATE

In the next update of the package in WP 4.1.6 we will include additional functionality which enables *finnsurveytext* to integrate with the popular *survey* package through the *svydesign* object. This will improve our package's useability in survey analysis and enable its analysis of open-ended questions to be better integrated with analysis of closed questions.

Additionally, the new version will upgrade comparison functions and visualisations to allow users to incorporate survey techniques such as weighting within analysis.

In Finland, it is typical that the frame population of a survey sample is formed by using register data. Some data from registers can also be included in the actual survey. The changes to the *finnsurveytext* package included in this update enable the integration of package results with this register data. As textual data have rarely been fully utilized in survey analysis, this provides an opportunity to analyse this text data while simultaneously taking into account the sampling design which allows the results obtained to be generalized beyond the survey.

Weighted comparison plot of most frequent words in reponses by respondents by region

ILLUSTRATIONS

The *finnsurveytext* package comes with comprehensive examples and sample data which guide users – including those who are inexperienced with the R programming language – to get started with analysing open-ended survey questions in Finnish.

The examples in this poster come from responses to question 11.3, 'Jatka lausetta: Maailman kolme suurinta ongelmaa ovat… ' ('Continue the following sentence: The three biggest problems in the world are… ') within the survey 'Young People's Views on Development Cooperation 2012' which is freely available for researchers from the Finnish Social Science Data Archive.

REFERENCE

Finnish Children and Youth Foundation: Young People's Views on Development Cooperation 2012 [dataset]. Version 2.0 (2019-01-22). Finnish Social Science Data Archive [distributor]. https://urn.fi/urn:nbn:fi:fsd:T-FSD2821