Knowledge-Driven Intelligent Survey Systems Towards Open Science

Elspeth Edelstein[1] · Jeff Z. Pan[2] · Ricardo Soares[2] · Adam Wyner[3]

Abstract

In this paper, we propose Knowledge Graph (KG), an articulated underlying semantic structure, as a semantic bridge between humans, systems, and scientific knowledge. To illustrate our proposal, we focus on KG-based intelligent survey systems. In state-of-the-art systems, information is hard-coded or implicit, making it hard for researchers to reuse, customise, link, or transmit structured knowledge. Furthermore, such systems do not facilitate dynamic interaction based on semantic structure. We design and implement a knowledge-driven intelligent survey system which is based on knowledge graph, a widely used technology that facilitates sharing and querying hypotheses, survey content, results, and analyses. The approach is developed, implemented, and tested in the field of Linguistics. Syntacticians and morphologists develop theories of grammar of natural languages. To evaluate theories, they seek intuitive grammaticality (well-formedness) judgments from native speakers, which either support hypotheses or provide counter-evidence. Our preliminary experiments show that a knowledge graph-based linguistic survey can provide more nuanced results than the traditional document-based grammaticality judgment surveys by allowing for tagging and manipulation of specific linguistic variables.

✉ Jeff Z. Pan
jeff.z.pan@abdn.ac.uk

1   School of Language, Literature, Music and Visual Culture, University of Aberdeen, Aberdeen, UK

2   Department of Computing Science, University of Aberdeen, Aberdeen, UK

3   School of Law and Department of Computer Science, Swansea University, Swansea, UK

Ohmsha ◼◼◼ ⌂ Springer

Introduction

In the present-day scientific work, researchers must collect and process large amounts of data. Involvement of human participants can make this task an especially complex one. Traditional document-oriented workflows are inherently unsuited to dealing with this kind of information, as new research is often presented in a fixed form that does not allow it to be re-purposed for further research or reproduced. The necessity to evolve such methods has become especially relevant with the emergence of Open Science, which emphasises transparency, re-usability, and knowledge sharing, encouraging flow of information through the scientific community. This sharing of knowledge accelerates the development of novel research, and allows the validation and expansion of previously tested hypotheses. In the era of Artificial Intelligence, these notions also apply to many other scenarios requiring human–system interaction, where a semantic bridge between humans and systems is needed.

In this paper, we propose Knowledge Graph (KG) [1, 2], an articulated underlying semantic structure, as such a semantic bridge between humans, systems, and information. To illustrate our proposal, we focus on KG-based intelligent survey systems. The advantages of the approach are that surveys are generated using the semantic information in the structure, the participants populate the structure, and survey interactions are based around specific semantic components. In addition, the approach facilitates transparency, transmission, and re-usability.

A popular approach to information gathering is *Amazon's Mechanical Turk* (MTurk),[1] a crowdsourcing platform where on-demand users do *Human Intelligence Tasks*, such as the completion of advertised surveys for a price. Another tool is SurveyMonkey,[2] which allows users to develop a survey online, present it to a community, and analyse results; surveys might have if-then-else structures. More relevant to our domain, there are online Linguistic experiments, which query users for linguistic judgments.[3] As useful as these tools are, information is often hard-coded or implicit in these systems, making it hard for researchers to reuse, customise, link, or transmit the knowledge. Furthermore, such systems do not easily facilitate dynamic interaction with the participant.

In our work, we develop and deploy a novel approach to survey-based practice by building in a survey system that uses Knowledge Graphs as an articulated underlying semantic structure, and which provides three different components of exposure to relevant levels of user: the participant in the survey, who answers the questions; the domain expert, who customises the knowledge structure to suit the problem; and the knowledge engineer, who constructs the underlying semantic structure. These will be discussed further below.

To test our survey system, we focus on an issue in Linguistics as specified by a linguist, who provides the domain knowledge. The tool represents linguistic information about the features and syntactic relationships in sentences. The user's task

---

in the survey is to judge a sentence acceptable or unacceptable. Given the survey results, the linguist has detailed information about the significant linguistic features and syntactic relationships. In addition, the linguist can incorporate alternative hypotheses, which are dependent patterns of features and syntactic relationships, into the system, allowing data gathering to test the alternatives. By enabling exploration of hypotheses and analysis of results into relevant components, the survey tool is a novel way to gather and analyse data.

As far as we know, this is the first effort to make use of designing an intelligent survey system based on knowledge graphs. It makes three contributions. First, it enriches the existing survey systems with Knowledge Graph, while hiding technical detail from both survey users and researchers. Second, it illustrates an example of knowledge graph-driven software engineering [3], offering a built-in semantic bridge for humans, systems, and information. And finally, it facilitates Knowledge Graph-driven research management in Open Science, wherein researchers can use structured information to share knowledge and data.

The rest of the paper is organized as follows. In the section "Vision: enabling open science via knowledge graphs", we present our vision of enabling Open Science by building research infrastructures based on knowledge graphs. In the section "Background", we briefly introduce the notion of knowledge graph, some basic ideas of using knowledge to facilitate scientific research, and the linguistic task of grammaticality judgment. In the section "Requirements analysis", we outline the core requirements that we consider in our knowledge-driven survey systems. The section "Design of knowledge graphs and system" introduces the design of the knowledge graph for our topic as well as the implementation of our intelligent survey system. In the section "Implementation", we outline the implementation and its evaluation. Relevant existing works are reviewed in the section "Related work". Finally, we conclude with some observations and outlooks in the section "Conclusion and outlook".

Vision: Enabling Open Science via Knowledge Graphs

When it comes to Open Science, many would think about open research data, but Open Science actually has much more to offer. Fecher and Friesike [4] propose five Open Science schools of thought:

1. the *democratic school*, on making research products, such as research data, available;
2. the *pragmatic school*, on making research more efficient by opening the scientific value chain and allowing collaborations;
3. the *infrastructure school*, on providing digital research infrastructure for research life-cycles;
4. the *public school*, on making science more understandable to the public and involving the public in the research process;
5. the *measurement school*, on alternative and faster impact measurement than impact factor.

Transparency, openness, and reproducibility are widely recognised as key features of science. There is a growing body of evidence [5], suggesting that transparency, openness, and reproducibility are not yet routine in daily scientific practice. Some might think that this could be due to the lack of an academic reward system that sufficiently incentivises Open Science. This might be true, but the problem is much bigger than that. As we argued in the Introduction, the traditional document-oriented scientific workflow is no longer fit for today's pace of scientific advances. Instead, scientific entities, such as research questions, hypotheses, experiments and observations, and their relations, should be first class citizens in scientific workflows. A well-known problem is the problem of "Knowledge Burying" identified by Mon [6]: under the current document-oriented workflow, all the interconnected scientific entities, including research questions, hypothesis, test data, experiments and observations, as well as their relations are buried in papers, making them hard to automatically extract and reuse. To make such knowledge accessible, organisations have to pay people or to use text-mining tools to extract them from papers. Unfortunately, the former approach is not scalable, while the latter one is far from perfect and thus results in a loss of knowledge.

Our vision is that scientific knowledge graphs, which are collections of interconnected science-related entities, can play a key role in realising the vision of Open Science. Indeed, scientific papers are essentially research products produced in research life-cycles, while scientific knowledge graphs are always there within the whole of research life-cycles.

In this paper, we are mainly focus on the infrastructure dimension and pragmatic dimension under Fecher and Friesike's classification. There are three stages in research life-cycles [7, 8]: in the *early stage*, researchers try to understand target research questions and to identify gaps from the existing work; in the *tentative stage*, researchers collect data, generate hypotheses from research questions, perform surveys and/or experiments, and record and analyse results from surveys and experiments; in the *finalised stage*, researchers produce reports and papers based on surveys and experiments, which will then be reviewed by the scientific community.

In this paper, we are not aiming at providing an infrastructure for all these three stages. Instead, we start to address the tentative stage from among the three stages of research life-cycles, as it is the core stage of the research life-cycle. From the incentive point of view, this could help fellow researchers to experience and understand the benefits of knowledge graph empowered research infrastructures.

Within the tentative stage, we select a survey as our first step. In other words, supporting survey is the focus of this paper. In research using human participants, a survey consists of a list of questions aimed at extracting specific data from a target group of people. Scientific surveys are used to increase knowledge in fields such as social science research.

In the next section, we will briefly introduce the notion of Knowledge Graph and a survey-based research topic in linguistic research, which we will build our case studies upon.

Background

Knowledge Graph

Knowledge Graph has become popular in knowledge representation and knowledge management applications widely across search engine, biomedical, media, and industrial domains [1]. In 2012, Google coined the term Knowledge Graph (KG) with a blog post-titled 'Introducing the Knowledge Graph: things, not strings'. In 2012, Knowledge Graph was added to Google's search engine in 2012 and a 'Knowledge Panel' is added to the search results page. Since then, Knowledge Graph has been widely used in the world's leading IT companies, not only on semantic search, but also on data integration, recommendation systems, as well as many intelligent applications.

Formally, a knowledge graph $\mathcal{G} = (\mathcal{D}, \mathcal{S})$ consists of a data sub-graph $\mathcal{D}$ of interconnected typed entities and their attributes, as well as a schema sub-graph $\mathcal{S}$ that defines the vocabulary used to annotate entities and their properties in $\mathcal{D}$. Facts in $\mathcal{D}$ are represented as triples of the following two forms:

- *property assertion* (h, r, t), where h is the head entity, r is the property, and t is the tail entity; e.g., (ACMilan, playInLeague, ItalianLeague) is a property assertion.
- *class assertion* (e, rdf:type, C), where e is an entity, rdf:type is the instance of relation from the standard W3C RDF specification, and C is a class; e.g., (ACMilan, rdf:type, FootballClub) is a class assertion.

A scheme sub-graph $\mathcal{S}$ includes Class Inclusion axioms $C \sqsubseteq D$, where C and D are class descriptions, such as the following ones: $\top \mid \bot \mid A \mid \neg C \mid C \sqcap D \mid \exists r.C \mid \leq n\ r \mid = n\ r \mid \geq n\ r$, where $\top$ is the top class (representing all entities), $\bot$ is the bottom class (representing an empty set), A is a named class r, r is a property, and n is a positive integer. For example, the types of River and City being disjoint can be represented as $River \sqsubseteq \neg City$, or $River \sqcap City \sqsubseteq \bot$. The W3C standard for defining the schema of knowledge graphs is OWL (Web Ontology Language), which is based on Description Logics [9]. The W3C standard for querying knowledge graphs is SPARQL. We refer the reader to [1] for a more detailed introduction of knowledge graphs.

The Linguistic Issue: Grammaticality Judgments

In investigating syntactic phenomena, linguists require data on what is judged *grammatical* by native speakers, i.e., what syntactic forms which they can and cannot use. This information may be obtained by asking speakers to provide *grammaticality judgments*, assessments of whether particular syntactic constructions are acceptable. Data of this type allow linguists to describe and define the parameters of natural language grammar as it is used. As such, native speaker

judgments of grammaticality are especially important in the study of 'non-Standard' sentence forms which differ from a more widely used 'Standard' norm, allowing researchers to establish the extent of syntactic variation within a language.

In a traditional *grammaticality judgment task*, a native speaker *participant* is presented with a series of sentences, which they rate on a scale of acceptability defined by the *linguist*. Although linguists often seek to measure the effects of specific linguistic features or variables, judgments are made at sentence level, meaning that the reasons for speakers' judgments may be obtuse to the researcher. For instance, a speaker might reject a sentence such as *My hair wants cut*, because:

- They require *to be* in a sentence such as this:
  My hair wants to be cut.
- They permit this syntactic construction, but only with the main verb *need*:
  My hair needs cut.
- They require an animate subject with the verb *want*:
  My cat wants fed.
- They require the verb *like* to be followed by a progressive participle:
  My hair wants cutting.

In a face-to-face interview, a linguist may use follow-up discussion to determine whether the participant has rejected the sentence for these reasons or others. However, this type of confirmation is not practical with large numbers of participants or surveys conducted online.

The specific variables of interest to the linguist may also be obscure to the participant, both because researchers may wish to conceal their exact object of study to preclude participants' knowledge biasing results, and because speakers cannot always access the implicit knowledge of language that underpins their judgments. Thus, even if questioned by the researcher, a participant may not articulate their reasoning, beyond stating that a sentence 'sounds wrong'.

As a test case, we investigate a syntactic construction found in Scottish and Northern Irish English, namely the use of the verbs *need*, *want*, or *like* followed directly by a passive participle. These constructions contrast with the 'Standard' use, where an auxiliary *to be* is present following the main verb.

– *The cat needs fed* (non-Standard: Scottish and Northern Irish English)
– *The cat needs to be fed* (Standard English)

A number of linguistic features may affect the use of the non-Standard form, especially for speakers who also allow the contrasting Standard construction. These features include the choice of the main verb (*need*, *want*, or *like*); whether the subject is animate (living and sentient) or inanimate; and whether the subject is definite (specific and known) or indefinite.

The above pair of sentences represents use of the main verb *need* with an animate, definite subject (*the cat*). They differ in the presence or absence of *to be*, which also constitutes a variable linguistic feature.

According to previous work on the non-Standard form, *need* is the most widely used main verb with this construction, followed by *want* and then *like* [10]. Inanimate subjects may also be more frequently used with *want* and *like* in the non-Standard form than in the Standard equivalent, suggesting that these verbs are semantically distinct from their Standard counterparts [11]. If true, these differences indicate that the non-Standard construction is syntactically distinct from the Standard form, rather than simply representing the Standard *to be* not being pronounced. We might also expect that the Standard form is acceptable to more speakers than the non-Standard form, although the reverse may be true for certain populations.

In our test case, *participants* were given a binary choice, mapped to the values of 0 (for *this sentence sounds strange to me*) and 1 (for *this sentence sounds good to me*).

Requirements' Analysis

In this section, we present the requirements for our knowledge-driven survey system. There are three sources of requirements: from the perspective of survey system in general, from the perspective of target research field(s) (linguistic domain in our case), and from the perspective of knowledge graph design. These requirements will be revisited in the evaluation of our case studies.

Scientific Survey System Requirements These requirements constitute the skeleton of what should be expected from any survey system, representing the most basic, yet essential functions.

- SR1: The researcher should be able to input data to the Knowledge Graph or modify the Knowledge Graph while creating surveys, without having to understand the notion of Knowledge Graph.
- SR2: The respondent should be able to access and respond to stimuli.
- SR3: The researcher should be able to query simple and complex patterns of results with respect to the Knowledge Graph structure.
- SR4: The researcher should be provided with statistical evaluation with respect to the Knowledge Graph.

Linguist Domain Requirements These are what the linguist needs for their task.

- LR1: The researcher should be able to input survey sentences and perform linguistic variable tagging on them.
- LR2: The respondent should be able to read sentences and input grammaticality or acceptability judgments
- LR3: The researcher should be able to analyse grammaticality judgments with respect to linguistic variable tags.
- LR4: The researcher should be able to test different hypothesis patterns in relation to single and multiple linguistic variables.
- LR5: The researcher should be able to obtain fine-grained results at both sentence and linguistic variable level.

Knowledge Graph Requirements

To make the system reusable to other subjects than Linguistics, we need to separate the basic concepts in generic survey systems from linguistic survey systems.

- KR1: The survey system knowledge graph should cover basic concepts related to the survey system.
- KR2: The linguistic feature knowledge graph should cover basic concepts needed in the linguistic surveys.

Design of Knowledge Graphs and System

According to the requirements, we need to have two knowledge graphs for the knowledge-driven survey system: one for generic survey systems and the other for linguistic surveys. We first present the schemas of the two knowledge graphs before presenting some example triples of the two knowledge graphs in "Design of knowledge graph". We then present our approach and design in "Approach and system design".

Design of Knowledge Graph

Survey Ontology

The survey ontology is a general purpose survey ontology which can be extended to specific domains such as Linguistics (cf. the Linguistic Feature Ontology section).

First, we identify key classes and properties in the survey ontology. Key classes include SurveyQuestion, AnswerOption, SurveyAnswer, and Hypothesis, Participation, and User, while key properties include: hasAnswerOption (connecting SurveyQuestions and AnswerOptions), hasAnswer (connecting Participation and SurveyAnswers), hasUser (connecting Participation and User), hasSurveyQuestion (connecting Participation and SurveyQuestion), and AnswerOptions), hasContent (connecting a survey question with its content to be defined in the domain-specific ontology). Note that we use the Participation class to represent the 3-ary relation among User, SurveyQuestion, and SurveyAnswer.

Second, we will need to specify the dependencies of the classes and properties in the survey ontology[4]:

- SurveyQuestion $\sqsubseteq\geq 1$ hasAnswerOption.AnswerOption (Each survey question has at least 1 answer option);
- SurveyQuestion $\sqsubseteq =1$ hasSurveyContent (Each survey question has exactly 1 content);
- Participation $\sqsubseteq =1$ hasSurveyUser.User (Each participation has exactly 1 user);

---

[4] To save space, we do not include domain and range axioms here.

- Participation $\sqsubseteq$ =1 hasSurveyQuestion.SurveyQuestion (Each participation has exactly 1 survey question);
- Participation $\sqsubseteq$ =1 hasSurveyAnswer.SurveyAnswer (Each participation has exactly 1 survey answer).

Linguistic Feature Ontology

The survey ontology is extended with domain-specific linguistic features. First, we identify key classes and properties in the linguistic survey ontology. Key classes include Sentence, POS (Part of Speech), Word, and Feature, while key properties include: hasPOS (connecting Sentence and POS), hasWord (connecting POS and Word), hasFeature (connecting Hypothesis/POS and Feature), hasString (connecting Sentence/POS/Word with some strings), and relatedFeature (connecting features).

Second, we will need to specify the dependencies of the classes and properties in the survey ontology[5]:

- SurveyQuestion $\sqsubseteq$ =1 hasContent.Sentence (Each survey question has exactly 1 sentence);
- Sentence $\sqsubseteq$ $\geq$ 1 hasPOS.POS (Each sentence has at least 1 POS);
- POS $\sqsubseteq$ $\geq$ 1 hasWord.Word (Each POS has at least 1 Word);
- Hypothesis $\sqsubseteq$ $\geq$ 1 hasFeature.Feature (Each hypothesis has at least one feature);
- Sentence $\sqsubseteq$ $\geq$ 1 hasString (Each sentence has some string);
- POS $\sqsubseteq$ $\geq$ 1 hasString (Each POS has some string);
- Word $\sqsubseteq$ $\geq$ 1 hasString (Each word has some string).

Parts of the linguistic feature ontology are constructed by linguistic researchers: (1) by providing a list of sub-classes of Feature, such as Subject or MainVerb (Subject $\sqsubseteq$ Feature, MainVerb $\sqsubseteq$ Feature), (2) using these sub-classes of Feature to annotate POSs in survey sentences (cf. next section).

Data Sub-graph example

In general, there are many linguistic linked data resources [12] online. To illustrate the two knowledge graphs, we consider an example survey sentence: *The cat needs fed*. For each sentence, there are two answer options: *Grammatical* and *Not grammatical*. Here are some triples related to this sentence:

- (Q1, hasContent, S1): the survey question Q1 has the sentence S1 as the content;
- (S1, hasString, '*The cat needs fed.*');
- (P1, rdf:type, POS): P1 is a POS;
- (S1, hasPOS, P1): S1 has a POS P1;
- (P1, hasString, 'The cat');
- (P1, rdf:type, Subject): P1 is annotated as an instance of Subject.

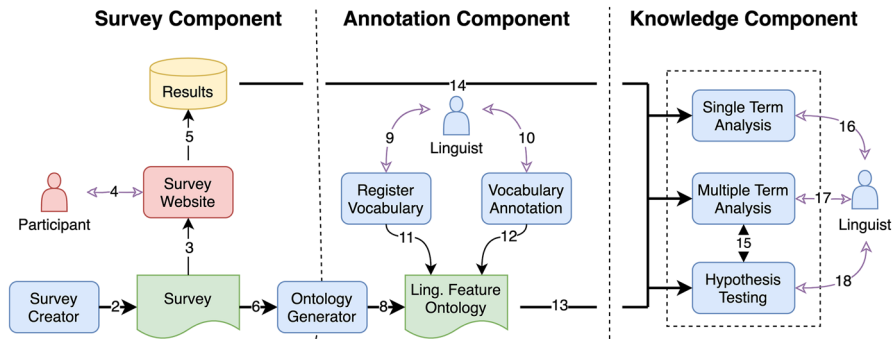[5] To save space, we do not include domain and range axioms here.

Fig. 1 Architecture diagram

The knowledge graphs serve as a bridge between researchers, the survey system, and the information in terms of understanding the sentences, survey answers, as well as related features. This could help e.g. semantically search for [13] related survey questions.

Approach and System Design

To provide a successful semantic-enabled Survey System to be of use to researchers who have no KG background, it is vital that the complexity of the system be obscured from them without sacrificing the leverage provided by the KG itself; linguists are users of an ontology and not experts in ontology management. In other words, the key challenges for the design of the knowledge-driven survey system are: (C1) how to embed knowledge graphs into a survey system, so that knowledge graphs serve as a bridge between the system, human researcher, and the information; (C2) how to do this in a transparent way, so that even the researchers who do not have a deep understanding of knowledge graph could use the system.

The challenge C2 suggests that the user interface should look similar to those of the existing systems, so that users can use it without a learning curve. We call such user interface component the *Survey Component*. Challenge C1 indicates that there should be some component dealing with the mapping between elements of the Survey Component and the knowledge graphs; we call this the *Annotation Component*. Finally, we have the *Knowledge Component* to exploit knowledge graphs to provide intelligent survey services.

In what follows, we will describe these three components in detail. Figure 1 presents the architecture diagram of the three components.

Survey Component

As shown in Fig. 1, the main processes that compose the *Survey Layer* are the *Survey Creator* and the *Survey Website*. It incorporates the functionalities of a survey without any explicit knowledge. The Researcher creates the survey that is presented to the *Participant*, and the *Participant* only interacts with the survey system at this
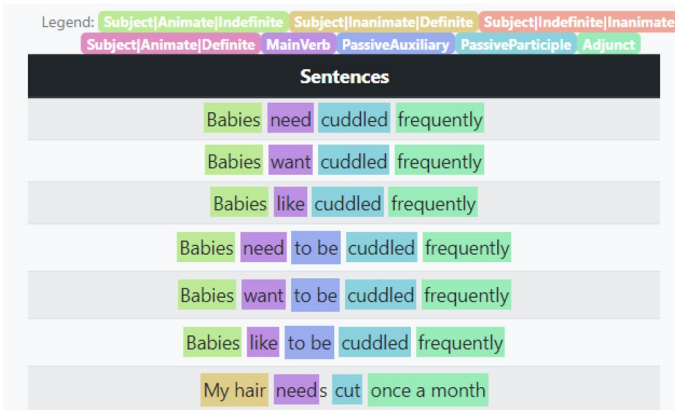
Fig. 2 Example of the annotations depicted by Linguists

component. The researcher is provided with an access link, which sent to *Participants* to complete the survey. Our platform stores the *Participants'* answers on its completion. The researcher can then explore the *Survey Results*. Theoretically, existing survey systems could potentially be reused as a survey component in our architecture.

Annotation Component

The main tasks of the Annotation Component include (AC1) maintaining the vocabulary (also known as terms) as *Features* in the *Linguistic Feature Ontology* and (AC2) annotating *POSs* in *Sentences* with the vocabulary (terms).

For the task of Vocabulary Registration (AC1), a user interface is needed for the researcher to add, update, and remove vocabulary, including *Features* and relations. New feature vocabulary proposed by the researcher can be added as sub-classes of the *Feature* class in the *Linguistic Feature Ontology*. Similarly, the new relation vocabulary will be added as sub-properties of the *relatedFeature* in the *Linguistic Feature Ontology*.

For the task of Sentence Annotation (AC2), another user interface (Fig. 2) is needed for the researcher to annotate the *Sentences* as she sees fit with the feature and relation vocabulary. For example, given the *Sentence The cat needs fed.*, the researcher can highlight part of the *Sentence*, such as *The cat*, and then annotate it with a feature vocabulary *Subject*. Some triples will be added into the *Linguistic Feature Ontology*, as discussed in the Linguistic Feature Ontology section.

Knowledge Component

The main task for the Knowledge Component is to provide intelligent survey services based on knowledge graphs, including *Single Term Analysis, Multiple Term Analysis* and *Hypothesis Testing*. 'Term' here refers to feature. Thus, single term analysis uses only one feature, while multiple term analysis uses more than one
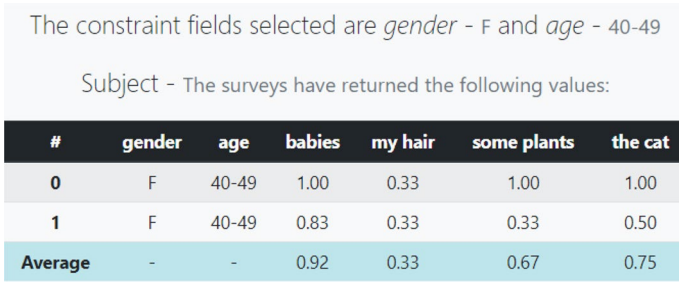
#	gender	age	babies	my hair	some plants	the cat
The constraint fields selected are *gender* - F **and** *age* - 40-49						
Subject - The surveys have returned the following values:						
0	F	40-49	1.00	0.33	1.00	1.00
1	F	40-49	0.83	0.33	0.33	0.50
Average	-	-	0.92	0.33	0.67	0.75

Fig. 3 Single term analysis of *Subject* with two constraints

feature. Hypotheses can be defined on top of multiple term analysis. All these three types of survey services are based on feature vocabulary.

Single Term Analysis This service is for the researcher to select a feature vocabulary to construct a single term query. Formally, given k sentences, n participants, and the feature *term*, a single term query $Q_S(term)$ is calculated as follows:

$$\frac{\sum_{j=0}^{n} \left(\frac{\sum_{i=0}^{k} score_{ij}*appear_i(term)}{count(term)} \right)}{n}, \tag{1}$$

where $score_{ij}$ is the score that participant j provided for sentence i, $count(term)$ is the total number of sentences containing instances of the feature *term*, and $appear_i(term)$ is 1 if some instance of the *term* appears in sentence i, otherwise 0.

Constraints can be added into single term queries. Typically, a constraint is applied on a field related to user related information, such as gender, age, or location. For example, Fig. 3 illustrates a single term query $Q_S(Subject)$ with $gender = Female, 40 \leq age \leq 49$ as the constraints. The result of the query is a table, the columns of which include the two constraints, as well as all the instances of the feature *Subject*.

In case there is only one instance of the feature *term*, we also compute $Q_S(\sim term)$, where $\sim$ is the Negation as Failure operator, meaning that we are looking for sentences that do not contain any instance of the given *term*. We combine the results of the two single term queries together for more insightful comparisons.

Multiple Term Analysis. This service is similar to the previous one, but with multiple terms. Formally, given k sentences, n participants, and the set of feature vocabulary $term(0), \dots, term(m)$, a multiple term query $Q_M(term(0), \dots, term(m))$ is calculated as follows:

$$\frac{\sum_{j=0}^{n} \left(\frac{\sum_{i=0}^{k} \left(score_{ij}*\prod_{t=0}^{T} appear_i(term(t)) \right)}{count(term(1),\dots,term(m))} \right)}{n}, \tag{2}$$

#	need+to be	need+Without to be	want+to be	want+Without to be	like+to be	like+Without to be
	MainVerb, PassiveAuxiliary – The surveys have returned the following values:					
0	1.00	1.00	0.75	0.75	0.75	0.75
1	1.00	1.00	0.50	0.50	0.50	0.00
2	1.00	1.00	0.25	0.00	0.50	0.25
3	0.75	0.75	0.75	0.75	0.75	0.25
4	1.00	1.00	0.75	0.75	0.75	0.50
5	1.00	0.25	1.00	0.50	0.75	0.25
6	1.00	1.00	0.50	0.50	0.75	0.75
7	1.00	0.00	0.75	0.00	1.00	0.25
Average	0.97	0.75	0.66	0.47	0.72	0.38

Fig. 4 Multiple Term Query example: *MainVerb* ('need', 'want', 'like') and *PassiveAuxiliary* ('to be')

where $score_{ij}$ is the score that participant j provided for sentence i, $count(term(1), \ldots, term(m))$ is the total number of sentences containing instances of every single feature within $term(0), \ldots, term(m)$, and $appear_i(term(t))$ is 1 if some instance of the $term(t)$ appears in sentence i, otherwise 0. Figure 4 illustrates a multiple term query $Q_M(MainVerb, PassiveAuxiliary)$. Note that *PassiveAuxiliary* has only one instance 'to be', and thus, Negation as Failure is applied by adding some columns about 'Without to be'.

Hypothesis Testing This service is to help the researcher to assess and register some hypotheses into the system, which could help monitor in real time if the registered hypotheses are satisfied by the results from the participants. We consider two types of hypotheses patterns (HP1 and HP2). All hypotheses are based on multiple term queries.

(HP1) Threshold hypotheses: given a multiple term query Q_M with its two columns $MC1$ and $MC2$, and two threshold values $t1$ and $t2$, a threshold hypothesis is defined as: $H_T(MC1, MC2, t1, t2)) = \neg(average(MC1) > t1) \vee (average(MC2) > t2)$. Informally, it says if $MC1$ crosses threshold $t1$, then $MC2$ should cross threshold $t2$.

(HP2) Comparator hypotheses: given a multiple term query Q_M with its two columns $MC1$ and $MC2$, and a comparator $\prec \in \{\leq, =, \geq\}$, a comparator hypothesis is defined as $H_T(MC1, MC2, \prec) = average(MC1) \prec average(MC2)$. Informally, it says $MC1$ is less ($\leq$)/equally ($=$)/more ($\geq$) acceptable than $MC2$.

Implementation

We implemented a web-based prototype for the proposed *Knowledge-Driven Survey System* in *Javascript* and *PHP*. The first functionality that is available to the *Researcher* is the building of a new Survey, using a drag-and-drop form editor (cf.

Fig. 5 Drag-and-drop interface during Survey Creation

Fig. 5). This incorporates the functionalities of the *formBuilder*[6] library, which is a flexible, scalable tool for survey construction. After building the desired survey structure, a *JSON* file is generated and properly adapted to be able to be received by a different library. This library, *surveyJS*,[7] is a powerful survey tool which prepares the outlook of a survey from a structured *JSON* file.

In the implementation of the Annotation Component, we allow the researcher to define a few different sets of vocabulary, so that she can have alternatives before she decides which set to use. Also, we allow the researcher to choose between annotating once only or to all the sentences containing the exact highlighted phrase. This helps significantly reducing the time needed for the researcher to annotate the sentences in the survey.

For single term queries in the Knowledge Component, in case there is only one instance of the feature *term*, we also compute $Q_S(\sim term)$, where $\sim$ is the Negation as Failure operator, meaning that we are looking for sentences that do not contain any instance of the given *term*. We combine the results of the two single term queries together for more insightful comparisons.

As for hypotheses pattern HP2 in the Knowledge Component, it should be noted that, even if $MC1 > MC2$ is true, it does not mean that all participants agrees that. Therefore, our survey system also provides the numbers of participants that agree to each of the three cases ($MC1 > MC2$, $MC1 = MC2$ and $MC1 < MC2$).

In the next section, we will present two case studies on grammaticality judgments and use them to evaluate our survey system.

Evaluation with Case Studies: Grammaticality Judgments

Case Study 1

Experiment Setup

As described in the section "The linguistic issue: grammaticality judgments", our case study examined the non-Standard use, found in Scotland and Northern Ireland, of verbs such as *need*, *want*, or *like* followed directly by a passive participle, as compared to more Standard use of these verbs followed by an auxiliary *to be* and passive participle.

- *The cat wants fed* (non-Standard).
- *The cat wants to be fed* (Standard).

The survey was set up by a linguistic researcher with no KG background, who established a vocabulary of relevant linguistic variables for this construction: main verb (*need*, *want*, or *like*), subject (in)animacy, subject (in)definiteness, and presence/absence of *to be*. The researcher then input and annotated 24 sentences covering all possible combinations of these linguistic features. In this iteration of the survey, the same subjects were used for each combination of (in)animacy and (in)definiteness, so that only four subjects were used across all the sentences:

- *the cat* [+ animate, + definite]
- *babies* [+ animate, − definite]
- *my hair* [− animate, + definite]
- *some plants* [− animate, − definite].

Each subject was also paired with an appropriate verb to ensure that speakers would not reject sentences for reasons of semantic anomaly (e.g., *My hair needs to be watered*):

- *the cat-fed*
- *babies-cuddled*
- *my hair-cut*
- *some plants-watered*.

Twelve respondents were recruited by word of mouth. They completed this pilot survey online by rating the sentences using a binary scale. All were native speakers of English born in Scotland or Northern Ireland and currently resident in Aberdeenshire.

Results are available as a mean rating (between 0 and 1) for each of the survey sentences; each individual respondent's rating is also available. In addition, results can be calculated for specific variables that occur in more than one sentence, and for combinations of variables.

Hypothesis Testing

The survey system has allowed examination of several hypotheses in relation to the data obtained. Multi-term analysis of the current results tells us that when *to be* is absent, *need* has a higher global acceptance rate (0.90) than *want* (0.46), and *want* has a higher acceptance rate than *like* (0.31), as predicted by the previous work.

On an individual sentence level, both of the sentences below, with the main verb *like*, and an animate, definite subject (*my hair*), are rejected by all speakers:

- My hair likes cut once a month.
- My hair likes to be cut once a month.

Many speakers accept the inanimate, indefinite subject *some plants* with *like* regardless of whether *to be* is present (0.75) or absent (0.50):

- Some plants like to be watered every day.
- Some plants like watered every day.

The higher acceptance rate for the Standard form is surprising in this instance, as it contradicts the assertion in the previous work that inanimate subjects were more likely to be accepted with *like* (and *want*) in the Scottish form without *to be*.

Analysis of Results

As well as looking at the non-Standard construction on its own, we can do more general comparison of equivalent constructions with and without *to be*. Globally, the Standard *to be* form has a higher acceptance rate (0.71) than the Scottish form without *to be* (0.56). Individual comparison for *need*, *want*, and *like* with and without *to be* shows the same result for each main verb (i.e., the *to be* form has a higher acceptance rate), indicating that the overall result truly represents greater global use of the Standard *to be* form among our respondents, and is not down to a dispreference for the non-Standard construction with a particular verb.

The hypothesis testing and analysis of results is possible through use of manual calculation based on the averaging of mean acceptance rates for each sentence. The test survey has only one value for each combination of variables, making this approach relatively straightforward. For instance, there is only one sentence with an animate, definite subject, the main verb *like*, and no auxiliary *to be*:

- The cat likes fed twice a day.

While the findings of this survey are broadly in line with the previous work on this non-Standard syntactic construction [11], the small number of participants means that the results cannot be taken as strongly conclusive. Because each combination of values for animacy and definiteness is tied to a single subject, responses may also have been affected by respondents' views about the real-world properties of these entities; for instance, the higher acceptance of *some plants* than *my hair* with *like* in

both the non-Standard and Standard constructions raises the possibility that some participants view plants as capable of volition. These judgments, therefore, still provide limited information about the effect of individual variables such as animacy.

Manual calculations for testing the above hypotheses on the small data set of the test survey take about 20 min. Annotation of linguistic variables in the survey planning stage took 5–10 min. There is, therefore, a considerable benefit to researchers in terms of time saved, which is likely to increase with survey complexity. Moreover, integration of hypothesis testing in the survey system allows immediate updating of results as more participants are added. Identification and annotation of linguistic variables also create materials that can be reused for future surveys on similar linguistic constructions, thereby decreasing the time required for initial survey design and input.

Case Study 2

Experiment Setup

Our second case study also examined the non-Standard use of a verb followed directly by a passive participle, as compared to the more Standard from with *to be*.

Again, the survey was set up by a linguistic researcher with no KG background. The survey employed the same range of linguistic variables for this construction: main verb (*need*, *want*, or *like*), subject (in)animacy, subject (in)definiteness, and presence/absence of *to be*. In this instance, 12 animate and 12 inanimate subjects were selected; each of these could also be definite or indefinite (e.g., [+ definite] *the cat*, [− definite] *some cats*). As before, these subjects were paired with appropriate participle verbs.

Six versions of the survey were created using these 24 subjects. For each of the six versions, all 24 subjects were used, so that no two sentences within a single version of the survey would have the same subject–participle verb combination. The subjects used with other variables (e.g., the main verb *need* with *to be*) differed across the six versions of the survey, so that no combination of linguistic features was represented by the same sentence. For instance, the following sentences represent use of a [+ animate, + definite] subject for the main verb *want* without *to be*; each was employed in a different survey:

- That horse wants ridden
- That baby wants cuddled
- The cat wants fed
- The dog wants walked
- The cow wants milked
- That sheep wants shorn.

In total, 144 distinct sentences were used across the six versions of the survey. This representation of the same combination of linguistic features with different

subjects meant that unanticipated interpretations for specific entities (e.g., the possible belief that plants can have volition) were less likely to skew global results.

Fifty participants were recruited by word of mouth and through social media, and completed the survey online. Each version of the survey was completed by a minimum of six participants, and a maximum of 12. As before, all participants had grown up in Scotland or Northern Ireland. We obtained a total of 1200 judgments over 144 sentences, more than quadrupling the 288 judgments over 24 sentences for the original survey.

Hypothesis Testing

Multi-term analysis of the data in the second version of the survey confirms the findings of the original. When *to be* is absent, *need* has a higher global acceptance rate (0.67) than *want* (0.44), and *want* has a higher acceptance rate than *like* (0.17), as predicted by the previous research.

For the Standard *to be* form, *need* has a similar acceptance rate to non-Standard *need* without *to be* (St: 0.68, NS: 0.67). The *to be* form with *want* has a slightly lower global acceptance rate than the form without (St: 0.37, NS: 0.44), while the *to be* form with *like* has a much higher global acceptance rate than the form without (St: 0.39, NS: 0.17).

Global acceptance rates			
Non-standard	*need*	*want*	*like*
+ *to be*	0.67	0.44	0.17
Standard	*need*	*want*	*like*
− *to be*	0.68	0.37	0.39

For the Non-Standard form without *to be*, *need* had a global rating of 0.70 for animate subjects, and 0.64 for inanimate; *want* had a global rating of 0.55 for animate subjects and 0.32 for inanimate; and *like* had a global rating of 0.22 for animate subjects and 0.11 for inanimate.

Non-standard acceptance rates			
Animate subject	*need*	*want*	*like*
	0.70	0.55	0.22
Inanimate subject	*need*	*want*	*like*
	0.64	0.32	0.11

For the Standard *to be* form, *need* had a global rating of 0.63 for animate subjects, and 0.73 for inanimate; *want* had a global rating of 0.59 for animate subjects and 0.15 for inanimate; and *like* had a global rating of 0.57 for animate subjects and 0.21 for inanimate.

Standard acceptance rates			
Animate subject	*need*	*want*	*like*
	0.63	0.59	0.57
Inanimate subject	*need*	*want*	*like*
	0.73	0.15	0.21

Analysis of Results

As in the first-case study, results are in line with the previous findings that the non-Standard use of *need* without *to be* is higher than that of *want* without *to be*, which in turn is higher than use of *like* without *to be*. There was a small global preference for the Standard *to be* form over the non-Standard form without *to be* (St: 0.48, NS: 0.43). This difference appears to be attributable to the much higher acceptance of *like* with *to be* among our participants (St: 0.39, NS: 0.17).

Although acceptance of inanimate subjects for non-Standard use of *like* (0.11) is lower than for acceptance of animate subjects (0.22), this difference is much smaller than that between *like* with *to be* with inanimate subjects (0.21) and inanimate ones (0.57). This result, therefore, appears to support the hypothesis that inanimate subjects are more acceptable with *like* in the non-Standard construction, with the lower acceptance rate for inanimate subjects in the non-Standard form than in the Standard (NS: 0.11, St: 0.21) reflecting a lower overall acceptance rate for *like* without *to be*.

For *want* without *to be* acceptance of inanimate subjects is 0.32, compared to 0.55 for animate subjects. It is notable that Standard *want* with *to be* is similarly acceptable with animate subjects (0.59), but much less acceptable with inanimate ones (0.15). Again, this result appears to support the hypothesis that inanimate subjects are more acceptable with *want* in the non-Standard construction than in the Standard one.

Sentence-level analysis, however, introduces a level of caution with respect to these conclusions about animate subjects with non-Standard *like* and *want*. Of the 12 sentences used across the six surveys with an inanimate subject and *like*, only four were accepted as grammatical by any respondents:

- Some bicycles like polished (1/6 respondents)
- Many rooms like tidied (1/9)
- There windows like cleaned (5/9)
- My hair likes cut (7/12).

In contrast, all but one of the 12 sentences with a inanimate subject and non-Standard *want* was accepted by at least one participant. In this respect, the argument that inanimate subjects are acceptable with non-Standard *want* is stronger than that for the same argument with *like*. These differences between variable-level and sentence-level information highlight the importance of variable-level tagging and use of different words and sentences to represent the same variables. Notably, although the results here raise further questions, they present a more nuanced picture than those

in Case Study 1, where the outright rejection of some sentences also meant outright rejection of the combinations of variables which they represented.

Evaluation of Requirements

Case Study 2 met SR1 by providing an interface for the researcher to input linguistic data; more specifically, it met LR1 by allowing the linguistic researcher to input and annotate the sentences to be presented to survey users. To our knowledge, the existing survey systems do not allow annotation of variables in this way, meaning that linguistic surveys permit only information on the sentence level to be incorporated during survey creation. Reuse of variable tags input into the system during Case Study 1 meant additional time saving in the implementation of Case Study 2.

Dissemination of the survey satisfied SR2 and LR2, with respondents able to access and complete the survey online. No technical knowledge was required, and participants were able to make grammaticality judgments with only written instructions. This function does not differ significantly from the existing systems, although ours has the potential for greater flexibility in the presentation of sentences.

The hypothesis testing function of the survey system met the requirements for SR3 and SR4 by allowing the researcher to query the results of the grammaticality judgment survey in relation to individual variables and combinations of multiple variables. This function also fulfills LR3 and LR4. It represents the most important contribution of the current system. Other survey systems can provide some statistical analysis for individual questions (sentences in the linguistic case), but do not provide the capacity to analyse responses according to an array of features or variables. This type of analysis can be accomplished manually, but involves significant time expense: data must be transferred to a format in which it can be processed and annotated (e.g., a spreadsheet), and each calculation must be done individually.

Finally, within the system, it was also possible to view and compare statistics for individual sentences as well as variables, thereby meeting LR5. While sentence-level analysis is possible in other survey systems, it is the use of this in conjunction with variable-level analysis that adds power to the current system.

Knowledge Graph Evaluation

Although our approach mainly focuses in the Linguistic Feature Ontology, it can be revised for any kind of survey, given the underlying use of Knowledge Graph. Thus, it is worthwhile to apply general validation techniques to our Linguistic Feature Ontology, using the six dimensions of Ontology quality, as discussed by Poveda–Villalón [14]:

- Human understanding—how comprehensive is the ontology? The ontology uses well-known linguistic concepts, is small, and is sufficient.
- Logical consistency—is the reasoning consistent? The functionalities of the system have been exhaustingly tested. The OWL Ontology was implemented in Protegé 5.2.0 and tested with Pellet.

- Modeling issues—what is the quality of the modeling decisions? The Linguistic Feature Ontology suits the particular domain; as such, various semantic properties such as inverse relationships were not needed. Yet, this represents a modeling decision that could be reassessed.
- Ontology language specification—does the ontology comply with OWL standards? Our ontology's syntax is correct, which is supported by the implementation in Protegé.
- Real-world representation—how aligned is the ontology with the application domain. The Linguistic Feature Ontology was developed with the close interaction of the Linguist researchers, ensuring a model appropriate to the domain and as fulfilling requirement KR2.
- Semantic application—is the ontology aligned with the embedding software? The Ontology supports the platform's functionalities.

Some of these dimensions have established evaluation metrics, e.g., *Logical consistency* and *Ontology language specification*. Other dimensions would require the development of appropriate means to evaluate performance. Some dimensions are largely *human-centric*, such as *Human understanding*, *Modeling issues*, and *Real-world representation*. For each, the ontology would be interrogated by relevant experts to assess the validity and coverage of the existing model, along with proposing alternatives. To facilitate this, a series of qualitative questions would be developed for the experts to use as the basis of interrogation; for example, we might have questions for the dimensions, respectively, such as:

- *Are there additional or alternative linguistic concepts to work with on the data?*
- *Are the semantic properties appropriate and in the necessary relationships?*
- *Are there relevant linguistic data that has not be incorporated along with its linguistic analysis?*

As some of the responses will likely be subjective and variant across experts, some means to highlight relevant, scoped responses would be essential. For the final dimension of *Semantic application*, standard software engineering evaluation methods would be applied, e.g., unit testing.

Related Work

Intelligent Surveys

There have been attempts regarding dynamized survey systems, such as the *Dynamic Intelligent Survey Engine* (DISE) [15]. The survey platform DISE aims to implement functionalities with a focus on customers' preferences and uses a wide variety of data collection methods. As with our system, it implements a flexible approach to survey creation. In comparison to our system, the survey creation methodology is less intuitive, as the researcher builds its structure through an XML file, an approach that is successful, but after some learning curve. *DISE*'s focuses on data collection

methods for a consumer-oriented domain. Most importantly, it cannot reason with knowledge. By applying semantics, we can analyse survey results at level and complexity that *DISE* does not.

Linguistic Surveys

Grammaticality judgment surveys have been developed online for a considerable time, through tools that aim to facilitate researchers in the field of linguistics.

MiniJudge [16] attempts to complement the traditional methodology in grammaticality judgment experiments with the statistical analysis provided from modern practices. It focuses on "minimalist" experiments—small respondent groups and sets of sentences, quick surveys, and a few other constraints.[8] Although *MiniJudge* does not provide the benefits of reasoning services and is limited to two binary factors as does our approach, it has advantages in complex statistical analysis and level of research.

Other relevant tools include *WebExp* [17] and *IBEX* [18] ("Internet Based EXperiments"). *WebExp* is used in Psycholinguistics for reaction data, a feature that exploit; yet, it does not make use of a knowledge struture. *IBEX* focuses in grammaticality judgment in different tasks such as *FlashSentence*, which presents the sentence for limited time, or *DashedSentence*, presenting the sentence word-by-word or chunk-by-chunk. They do not encompass any novel analysis; in comparison of our system, they dwell entirely in the survey component, extending the capabilities of the original grammaticality judgment task.

MTurk Surveys

Two final tools are discussed, developed with the focus of running linguist-focused tasks with the crowdsourcing platform *Amazon Mechanical Turk*. The first, *Turkolizer* [19], takes a different, domain-specific approach to individual variables, using, just like the *MiniJudge* implementation, the concept of experimental factors (a simple example is provided by *Gibson* regarding two factors with two conditions each, where sentences are defined by *Subject-Object* order, and by having two or three question words. Each combination is mapped to a sentence, and since we have two binary factors, this would represent a four-sentence design). The last tool is called *Turktools* [20], inspired by *Turkolizer*, and it also implements its version of factorial design.

The surveys discussed do not provide the degree of freedom our knowledge-powered services offer through individual variables, as these systems hard-code the necessary variables upon survey creation. We provide a new depth of meaningful results, without big expense; a strength possible due to the Knowledge Graph that powers the present Survey System.

---

[8] see http://www.ccunix.ccu.edu.tw/~lngmyers/MJFAQ.htm.

Conclusion and Outlook

In this paper, we present an inviting approach to knowledge-driven survey systems, building a case for further interest and development on KG-driven software engineering [3], in Open Science and beyond. We also investigate a new solution for Grammaticality Judgment Tasks, proving the efficiency of our system by extending our Ontology to satisfy the Linguist researcher's needs. Our approach is a step forward in a field of study where knowledge graph technologies are not yet applied, presenting with our implementations the advantages that can be retrieved.

From the application front, there is much work that can be done, in surveys based on Psycholinguistics and other application domains. Even more intriguing approaches can be developed in this field by implementing further reasoning services, introducing the creation of properties and the disjoint sub-classes to the Linguist researcher, and the expansion of the Linguistic Feature Ontology to other relevant topics in Psycholinguistics. We also envision a modular approach which would allow our platform to extend to different application domains, linking disjoint areas semantically to our Survey Ontology. We also plan to make the survey more dynamic, so as to more closely simulate the level of flexibility allowed in face-to-face surveys. Last but not least, such dynamics scenarios might involve some uncertain information [21, 22] to be stored and used in knowledge graphs.

From the perspective of the infrastructure school of Open Science, future work will also include some knowledge graph-based research infrastructure for the other two stages, i.e., the *early stage* and *finalised stage*. The work is currently in the *tentative stage*, as it can be further interrogated as outlined in "Knowledge graph evaluation". Following the results of the evaluation, the analysis can be further, incrementally developed till the model and tool are stable, and we can work on report generation in a *finalised stage*.

References

1. Pan, J.Z., Vetere, G., Gomez-Perez, J. Manuel , Wu, H.: Exploiting Linked Data and Knowledge Graphs for Large Organisations. Springer, Berlin. ISBN 978-3-319-45654-6 (2016)

2. Pan, J.Z., Calvanese, D., Eiter, Th., Horrocks, I., Kifer, M., Lin, F., Zhao, Y.: Reasoning Web: Logical Foundation of Knowledge Graph Construction and Querying Answering. Springer, Berlin. ISBN 978-3-319-49493-7 (2017)
3. Pan, J.Z., Staab, S., Assmann, U., Ebert, J., Zhao, Y.: Ontology-Driven Software Development. Springer, Berlin. ISBN 978-3-642-31226-7 (2013)
4. Fecher, B., Friesike, S.: Open Science: One Term, Five Schools of Thought, Opening Science. Springer, Berlin (2014)
5. Nosek, B.A., et al.: Promoting an open research culture. Science **348**(6242), 1422–1425 (2015)
6. Mons, B.: Which gene did you mean? BMC Bioinform. **6**, 142 (2005)
7. Giunchiglia, F., Chenu-Abente, R.: Scientific Knowledge Objects, vol. 1. University of Trento, Trento (2009)
8. Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., Couch, P., Cruickshank, D., Delderfield, M., Dunlop, I., Gamble, M.: Why linked data is not enough for scientists. Future Gener. Comput. Syst. **29**(2), 599–611 (2013)
9. Baader, Franz, Calvanese, Diego, McGuinness, Deborah, Nardi, Daniele, Patel-Schneider, Peter F.: The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, Cambridge (2010)
10. Murray, T., Simon, B.L.: At the intersection of regional and social dialects: the case of like + past participle in American English. Am. Speech **77**(1), 32–69 (2002)
11. Edelstein, E.: This syntax needs studied. In: Zanuttini, R., Horn, L.R. (eds.) Micro-Syntactic Variation in North American English, pp. 242–268. Oxford University Press, Oxford (2014)
12. Chiarcos, C., Cimiano, P., Declerck, T., McCrae, J.P.: Linguistic linked open data. Introduction and overview. In: Proceedings of the 2nd Workshop on Linked Data in Linguistics (LDL-2013): Representing and Linking Lexicons, Terminologies and Other Language Data, pp. 1–9 (2013)
13. Pan, J.F., Taylor, S., Thomas, E.: Reducing ambiguity in tagging systems with folksonomy search expansion. In: Proc. of the 6th European Semantic Web Conference (ESWC2009) (2009)
14. Poveda-Villalón, M., Gómez-Pérez, A., Suárez-Figueroa, M.C.: Oops!(ontology pitfall scanner!): an on-line tool for ontology evaluation. J. Semant. Web Inf. Syst. **10**(2), 7–34 (2014)
15. Schlereth, C., Skiera, B.: DISE: Dynamic Intelligent Survey Engine. In: Diamantopoulos, A., Fritz, W., Hildebrandt, L. (eds.) Quantitative Marketing and Marketing Management, pp. 225–243. Gabler Verlag, Wiesbaden (2012)
16. Myers, J.: MiniJudge: Software for minimalist experimental syntax. In: Proceedings of the 18th Conference on Computational Linguistics and Speech Processing, pp. 271–285 (2006)
17. Keller, F., Gunasekharan, S., Mayo, N., Corley, M.: Timing accuracy of web experiments: a case study using the WebExp software package. Behav. Res. Methods **41**(1), 1–12 (2009)
18. Ibex 0.3.8 Manual. http://spellout.net/latest_ibex_manual.pdf. Accessed 1 Aug 2018
19. Gibson, E., Piantadosi, S., Fedorenko, K.: Using Mechanical Turk to obtain and analyze English acceptability judgments. Lang. Linguist. Compass **5**(8), 509–524 (2011)
20. Erlewine, M.Y., Kotek, H.: A streamlined approach to online linguistic surveys. Nat. Lang. Linguist. Theory **34**(2), 481–495 (2016)
21. Qi, G., Pan, J.Z., Ji, Q.: A possibilistic extension of description logics. In: Proc. of 2007 International Workshop on Description Logics (DL2007) (2007)
22. Pan, J.Z., Stoilos, G, Stamou, G.: Vassilis Tzouvaras and Ian Horrocks. f-SWRL: a fuzzy extension of SWRL (Extended version). J. Data Semant. **4090/2006**, 28–46 (2006)

Elspeth Edelstein is a Lecturer in Language and Linguistics at the University of Aberdeen. Her research focusses on English syntax and morphosyntactic variation between dialects.

Jeff Z. Pan received his Ph.D. in computer science from The University of Manchester. He is a Reader in the Department of Computing Science at University of Aberdeen and the Director of the Joint Research Lab on Knowledge Engineering and Information Security. His research focuses primarily on knowledge representation and artificial intelligence, in particular, on knowledge-based reasoning and learning, as

well as their applications. He is the Chief Editor of the first two books on Knowledge Graph. He is a PC Co-Chair International Semantic Web Conference (ISWC2020) and an Associate Editor of the Journal of Web Semantics (JWS).

Ricardo Soares graduated from his MSc in Artificial Intelligence from the University of Aberdeen. He is currently a Machine Learning Engineer at Fractal Labs, tackling new challenges in the FinTech space.

Adam Wyner has a PhD in Linguistics (Cornell University, 1994) and a PhD in Computer Science (King's College London). He is currently an Associate Professor at Swansea University, holding a joint position in the School of Law and Department of Computer Science. As an established researcher in Artificial Intelligence and Law, he published and presented on natural language processing (rule-based and machine learning), information extraction, ontologies, argumentation, controlled languages, case-based reasoning, policy consultations, and semantic web. He is on the OASIS Technical Committee for Legal-RuleML, which is developing a machine-readable standard for legal rules.