**Conference contribution :**

Soares, R., Edelstein, E., Pan, J. & Wyner, A. *Knowledge Driven Intelligent Survey Systems for Linguists.* Semantic Technology, -18).

http://dx.doi.org/10.1007/978-3-030-04284-4_1

# Knowledge Driven Intelligent Survey Systems for Linguists

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

[1] Department of Computing Science, University of Aberdeen, Aberdeen
[2] School of Language, Literature, Music & Visual Culture, University of Aberdeen, Aberdeen
[3] School of Law & Department of Computer Science, Swansea University, Swansea

**Abstract.** The paper introduces a novel approach to survey data collection that uses an articulated underlying semantic structure. In state of the art systems, knowledge is hard-coded or implicit in these systems, making it hard for researchers to reuse, customise, link, or transmit the structured knowledge. Furthermore, such systems do not facilitate dynamic interaction based on the 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 a theory or provide counter-evidence. Our preliminary experiments show that a knowledge graph based linguistic survey can provide more nuanced results than traditional document-based grammaticality judgment surveys by allowing for tagging and manipulation of specific linguistic variables.

**Keywords:** Knowledge Graph · Intelligent Survey System · Linguistic Acceptability Judgment Task.

## 1 Introduction

In this paper we present a novel approach to gathering information with online surveys using Knowledge Graphs as an articulated underlying semantic structure. 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) [4], a crowdsourcing platform where on-demand users do *Human Intelligence Tasks*, such as the completion of advertised surveys for a price. Another

[4] https://www.mturk.com/

tool is SurveyMonkey [5], which allows users to develop a survey online, serve 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.[6] As useful as these tools are, knowledge 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.

Knowledge Graphs, an approach to knowledge management and reasoning [1], have been increasingly applied to a range of different application domains, e.g. health-care *SNOMED-CT (Clinical Terms)* [3], marine species *MarineTLO* [4], and Linguistics *Linguistic Linked Open Data* [5], which attempts to link and integrate resources from various fields of Linguistics into the LLOD Cloud[7].

In our approach, 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 users: the participant in the survey, who answers the questions; the domain expert, who uses 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 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 existing survey systems with KG, while hiding technical detail from survey users and linguists. Second, it contributes to research on language variation, which opens opportunities to integrate variant languages into NLP tools. And finally, it facilitates knowledge-graph driven research management, wherein researchers can use structured information to share knowledge and data.

The rest of the paper is organized as follows. In Section 2, we briefly introduce the notion of knowledge graph, some basic idea of using knowledge to facilitate scientific research and the linguistic task of grammaticality judgment. In Section 3, we outline the core requirements that we consider in our knowledge-driven survey systems. Section 4 introduces our design of the knowledge graph for our topic as well as the implementation of our intelligent survey system. In Section

[5] https://www.surveymonkey.com

[6] `https://www.psytoolkit.org/`

[7] http://linguistic-lod.org/

5, we outline the implementation and its evaluation. Relevant existing works are reviewed in Section 6. Finally, we conclude with some observations and outlooks in Section 7.

## 2 Background

### 2.1 Knowledge Graph

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, and r the property and t 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 $\sqsubseteq$ $\bot$. We refer the reader to [2] for a more detailed introduction of knowledge graphs.

### 2.2 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 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 allows linguists to describe and define the parameters of natural language grammar as it is spoken. 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. Clarification may be provided through follow-up discussion with participants, but this solution is not practical with large numbers of participants or surveys conducted online. Moreover, the specific variables of interest to the linguist may be obscure to the participant.

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

- *The cat needs fed* (Scottish English)
- *The cat needs to be fed* (Standard English)

A number of linguistic features may affect the use of the Scottish 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 represent 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.

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*).

Previous work on the Scottish form indicates *need* is the most widely used main verb with this construction, followed by *want* and then *like* [6]. Inanimate subjects may also be more frequently used with *want* and *like* in the Scottish form than in the Standard equivalent [7]. We would also expect that the Standard form is acceptable to more speakers than the Scottish form, although the reverse may be true for certain populations.

## 3 Requirements Analysis

In this section, we present the requirements for our knowledge driven survey system. There are three categories of requirements - on the survey system, on the linguistic domain, and on the knowledge graph design.

**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 respondent should be able to read sentences and input judgments rating their grammaticality.
- SR2: The researcher should be able to query simple and complex patterns of results with respect to the Knowledge Graph structure.
- SR3: 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.
- 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 perform linguistic variable tagging on survey sentences.
- LR2: The researcher should be able to analyse grammaticality judgments with respect to linguistic variable tags.
- LR3: The researcher should be able to test different hypothesis patterns in relation to single and multiple linguistic variables.
- LR4: The researcher should be able to obtain fine-grained results at both sentence and linguistic variable level.

**Knowledge Graph requirements.** In order 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.

## 4 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, while the other for linguistic surveys. We firstly present the schemas of the two knowledge graphs before presenting some example triples of the two knowledge graphs in Section 4.1. We then present our approach and design in Section 4.2.

### 4.1 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. Section 4.2).

Firstly, we identify key classes and properties in the survey ontology. Key classes include SurveyQuestion, AnswerOption, SurveyAnswer and Hypothesis, Participation, 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.

Secondly, we will need to specify the dependencies of the classes and properties in the survey ontology:[8]

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

- 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);
- 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. Firstly, 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).

Secondly, we will need to specify the dependencies of the classes and properties in the survey ontology:[9]

- 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) by using these sub-classes of Feature to annotate POSs in survey sentences (cf. next section).

**Data Sub-Graph example** 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.*');

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

- (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.

The knowledge graphs serve as a bridge between researchers and the survey system in terms of understanding the sentences, survey answers, as well as related features.

### 4.2 Approach and System Design

To provide a successful semantic-enabled Survey System to be of use to researchers, 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 and human researchers; (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 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. Fig.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 component. The researcher is provided with an access link, which sent to *Participants* in order 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 including (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).

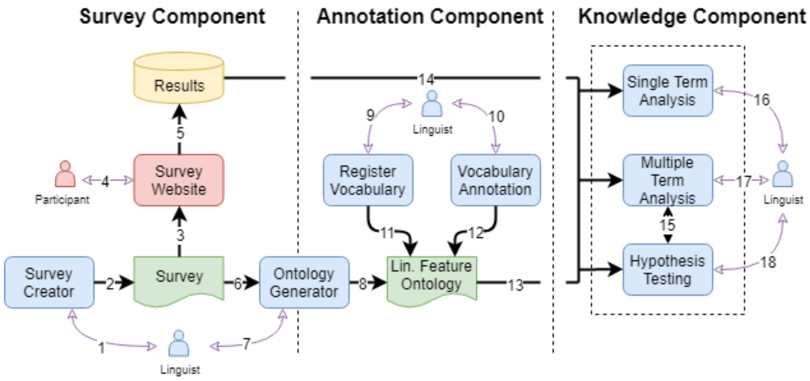
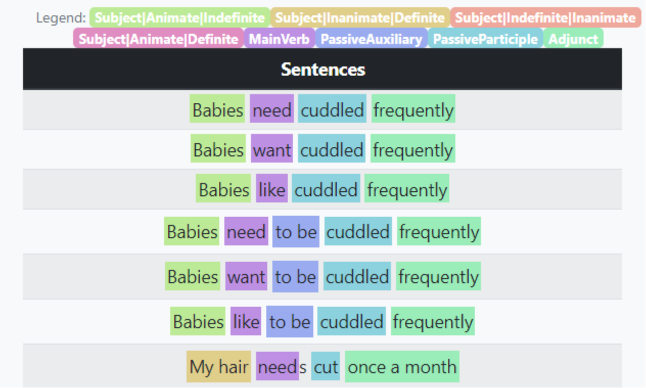


**Fig. 1.** Architecture Diagram

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 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 Section 4.3.


Legend: Subject|Animate|Indefinite Subject|Inanimate|Definite Subject|Indefinite|Inanimate Subject|Animate|Definite MainVerb PassiveAuxiliary PassiveParticiple Adjunct

| Sentences |
|---|
| Babies need cuddled frequently |
| Babies want cuddled frequently |
| Babies like cuddled frequently |
| Babies need to be cuddled frequently |
| Babies want to be cuddled frequently |
| Babies like to be cuddled frequently |
| My hair needs cut once a month |

**Fig. 2.** Example of the annotations depicted by Linguists.

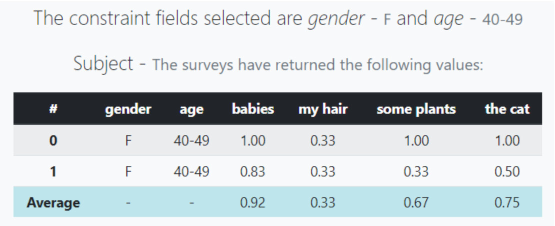
**Knowledge Component.** The main task for the Knowledge Component is to provide intelligent survey services based on knowledge graphs, including *Single Term Analysis, Multiple Term Analysis* and *Hypothesis Testing*. 'Term' here refers to feature. Thus single term analysis uses only one feature, while multiple term analysis uses more than one 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 $term$, $appear_i(term)$ 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 $Subject$.



The constraint fields selected are *gender* - F and *age* - 40-49

Subject - The surveys have returned the following values:

| # | gender | age | babies | my hair | some plants | the cat |
|---|---|---|---|---|---|---|
| 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 |

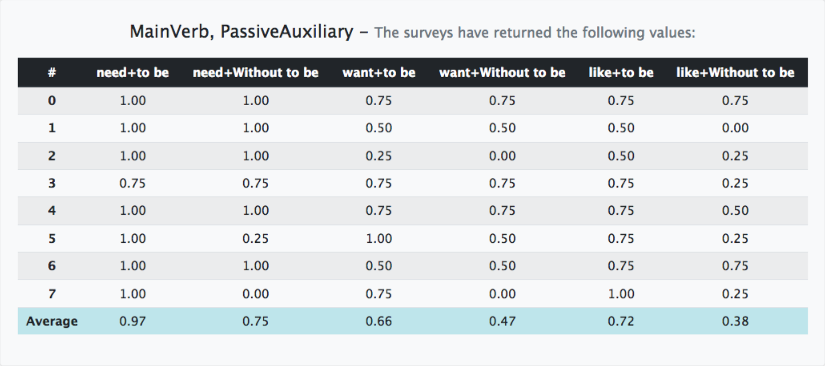
**Fig. 3.** Single Term Analysis of *Subject* with two constraints.

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 $term(0), ..., term(m)$, a multiple term query $Q_M(term(0), ..., term(m))$ is calculated as follows:

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

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



MainVerb, PassiveAuxiliary – The surveys have returned the following values:

| # | need+to be | need+Without to be | want+to be | want+Without to be | like+to be | like+Without to be |
|---|---|---|---|---|---|---|
| 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')

*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.

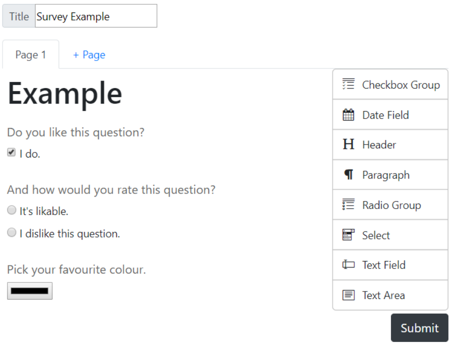
(HP1) Threshold hypotheses: given a multiple term query $Q_M$ with its two columns $MC1$ and $MC2$, and two threshold values $t1$ and $t2$, a threshold hypothesis is defined as: $H_T(MC1, MC2, t1, t2)) = \neg(average(MC1) > t1) \vee (average(MC2) > t2)$. Informally, it says if $MC1$ crosses threshold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

# 5 Implementation and Evaluation

## 5.1 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. This incorporates the functionalities of the *formBuilder*[10] 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*[11], is a powerful survey tool which prepares the outlook of a survey from a structured *JSON* file.



Title Survey Example

Example

Do you like this question?

☑ I do.

And how would you rate this question?

○ It's likable.

○ I dislike this question.

Pick your favourite colour.

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

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.

[10] https://formbuilder.online/

[11] https://surveyjs.io/

### 5.2 Case Study: Grammaticality Judgments

**Experiment Setup: Evaluating SR1 and LR1** As described in Section 2.2, our case study examined the use, found in Scotland, 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*
- *The cat wants to be fed*

The survey was set up by a linguistic researcher with no KG background, who established a vocabulary of relevant linguistic variables: 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.

Twelve respondents completed our pilot survey 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.

**Hypotheses Testing: Evaluating LR2, LR3, and LR4** 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 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 (.75) or absent (.50).

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

The higher acceptance rate for the Standard form is surprising in this instance, as it contradicts the assertion in 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: Evaluating SR2, SR3, and SR4** As well as looking at the Scottish 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, and is not down to a dispreference for the Scottish 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*

More comprehensive data collection would involve more sentences with the same variable values and more variables. For instance, given the much higher rate of acceptance for the subject *some plants* than *my hair*, other inanimate subjects would be needed in order to establish whether there is any specific pattern pertaining to (in)animacy. With only sentence-level statistics available, these additions would complicate the manual calculations required for hypothesis testing, requiring additional researcher time, and introducing more room for human error.

Manual calculations for testing the above hypotheses on the small data set of the test survey take about twenty minutes. Annotation of linguistic variables in the survey planning stage took 5-10 minutes. 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 creates materials that can be reused for future surveys on similar linguistic constructions, thereby decreasing the time required for initial survey design and input.

### 5.3 Knowledge Graph Evaluation

Although our approach mainly focuses in the Linguistic Feature Ontology, it can be revised for 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 [8]:

- 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.

## 6 Related Work

### 6.1 Intelligent Surveys

There have been attempts regarding dynamized survey systems, such as the *Dynamic Intelligent Survey Engine* (DISE) [11]. 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.

### 6.2 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* [12] 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 [12]. 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.

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

Other relevant tools include *WebExp* [14] and *IBEX* [15] ("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.

### 6.3 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* [16], 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 4-sentence design). The last tool is called *Turktools* [17], 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.

## 7 Conclusion and Outlook

In this paper, we present an inviting approach to Knowledge-driven survey systems, building a case for further interest and development. 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 needs. We believe our approach to be a step forward in a field of study where semantic technologies are not yet applied, presenting with our implementations the advantages that can be retrieved.

There is much work that can be done, in surveys based in 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.

## References

1. Pan, J.Z., Matentzoglu, N., Jay, C., Vigo, M., Zhao, Y.: Understanding Author Intentions: Test Driven Knowledge Graph Construction. In: Pan, J.Z. et al. (eds) Reasoning Web: Logical Foundation of Knowledge Graph Construction and Query Answering. Reasoning Web 2016. LNCS 9885, pp. 1-26. Springer, (2017)
2. Jeff Z. Pan, Guido Vetere, Jose Manuel Gomez-Perez, Honghan Wu. (Eds.). Exploiting Linked Data and Knowledge Graphs for Large Organisations. Springer. 2016.
3. Donnelly, K.: SNOMED-CT: The Advanced Terminology and Coding System for eHealth. Studies in health technology and informatics 121, 279-290 (2006)
4. Tzitzikas Y. et al.: Integrating Heterogeneous and Distributed Information about Marine Species through a Top Level Ontology. In: Garoufallou E., Greenberg J. (eds) Metadata and Semantics Research. MTSR 2013. Communications in Computer and Information Science, vol 390, pp 289-301. Springer, (2013)
5. 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)
6. Murray, T. and Simon, B. L..: At the intersection of regional and social dialects: The case of *like* + past participle in American English. *American Speech*, 77(1):32–69 (2002)
7. Edelstein, E.: This syntax needs studied. In Zanuttini, R. and Horn, L. R. (eds) Micro-syntactic variation in North American English, 242-268. Oxford University Press (2014)
8. Poveda-Villalón, M., Gómez-Pérez, A., Suárez-Figueroa, M.C.: Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation. Journal on Semantic Web and Information Systems, **10**(2), 7-34 (2014)
9. Giunchiglia, F. and ChenuAbente, R.: Scientific Knowledge Objects v. 1. University of Trento (2009)
10. 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 Generation Computer Systems, **29**(2), 599-611 (2013)
11. 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)
12. Myers, J.: MiniJudge: Software for minimalist experimental syntax. In: Proceedings of the 18th Conference on Computational Linguistics and Speech Processing, pp. 271-285 (2006)
13. Chen, T.Y., Myers, J.: Worldlikeness: A Web-based Tool for Typological Psycholinguistic Research. Univ. of Penn. Working Papers in Linguistics, **23**(1), 21-30 (2017)
14. Keller, F., Gunasekharan, S., Mayo, N., Corley, M.: Timing accuracy of web experiments: A case study using the WebExp software package. Behavior research methods, **41**(1), 1-12 (2009)
15. Ibex 0.3.8 Manual, http://spellout.net/latest_ibex_manual.pdf. Accessed 1/08/18
16. Gibson, E., Piantadosi, S., Fedorenko, K.: Using Mechanical Turk to obtain and analyze English acceptability judgments. Language and Linguistics Compass, **5**(8), 509-524 (2011)
17. Erlewine, M.Y., Kotek, H.: A streamlined approach to online linguistic surveys. Natural Language & Linguistic Theory, **34**(2) ,481-495 (2016)