A Dynamic and Informative Intelligent Survey System based on Knowledge Graph

Patrik Bansky[1], Elspeth Edelstein[2], Jeff Z. Pan[1][*], and Adam Wyner[3]

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

Abstract. In the paper we propose a dynamic and informative solution to an intelligent survey system that is based on knowledge graph. To illustrate our proposal, we focus on ordering the questions of the questionnaire component by their acceptance, along with conditional triggers that further customise participants' experience, making the system dynamic. Evaluation of the system shows that the dynamic component can be beneficial in terms of lowering the number of questions asked and improving the quality of data, allowing more informative data to be collected in a survey of equivalent length. Fine-grained analysis allows assessment of the interaction of specific variables, as well as of individual respondents rather than just global results. The paper explores and evaluates two algorithms for the presentation of survey questions, leading to additional insights about how to improve the system.

Keywords: Intelligent survey system · Dynamic and informative system · Linguistic grammaticality judgements

1 Introduction

This paper is about how to use knowledge graph to build an intelligent survey system. In fields such as Linguistics, Psychology, and Medicine, researchers rely on data from human participants, which are gathered either by verbal communication, written questionnaires, or Internet-based questionnaires. Online surveys are particularly popular in contemporary research due to their global reach, flexibility, ease of data analysis, and low administration cost, among other advantages [10]. However, research suggests that participant motivation in surveys decreases over time such that respondents are likely to engage in a sub-optimal way, lowering the overall quality of data collected [15]. Respondents may be reluctant to complete surveys due to low interest in participation, resulting in decreased response rates overall [18,28]. Internet-based questionnaires are also by nature less interactive than face-to-face data collection, limiting researchers' ability to follow up participants' answers in order to aid interpretation of results. In addition, while online questionnaires may be used to assess what populations find

---

[*] Corresponding author

acceptable and the inferences made from global percentages, it would be informative to analyse of each respondent's answers to uncover fine-grained patterns and (in)consistencies. That is, it would be worthwhile to measure the judgements of individual participants rather than generalising over groups.

To overcome these restrictions on quality and quantity of information, it is widely believed [5,2,12,27] that a more dynamic approach to questionnaires would lead to a higher standard of data collected. One important factor in developing questionnaires is the *order of the questions. Research methods in psychology* [8] argues that the most interesting questions should be ranked at the beginning of the survey to catch a respondent's attention, while less important questions should be near the end.

Our previous work [31] proposes an architecture of knowledge driven intelligent survey system. The idea is to use Knowledge Graph as a semantic bridge between humans and computational systems, so as to facilitate customisability, transmission, re-usability, explainability and extensibility. The system provides three different components of exposure to relevant users: (1) the participants of the survey, (2) the domain experts, and (3) the knowledge engineers. The participants simply answer the questions: their role is to judge whether sentences are acceptable or unacceptable. The domain experts, such as Linguists, customise the knowledge structure to fit their needs, and knowledge engineers construct the basis of the semantic structure. For example, a knowledge driven intelligent survey system on linguistics provides information about syntactic relationships and features of each sentence. Upon participant submissions, researchers are able to see detailed information about these features and syntactic relationships. In addition, using the data collected, researchers are able to use the tool to organise and analyse data using pattern features and syntactic relationships to either confirm or refute their original hypotheses.

In this paper, we further develop the notion of knowledge graph based dynamic survey system that responsively selects questions from a larger pool provided by the researcher. Prioritisation of questions is based on interaction of researcher hypotheses and participant input, allowing optimisation of data quality and user responses. Although the proposed survey system is built for linguistic judgements, for the sake of evaluation, such an explainable architecture can in principle be applied to many other scenarios, such as one for monitoring a local community's opinion on whether an existing nuclear power plant should resume its operations or not. To enable such domain adaptation, per-subject optimisations or randomisation should be in place, in particular when the target issue is sensitive in having manipulations from the researcher, such as the case of nuclear power plant.

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 [26,25] for a more detailed introduction of knowledge graphs.

2.2 Linguistic Background

Theoretical linguists working on morphosyntax, the structure of words and sentences, may use questionnaires to gather data needed to investigate grammatical structure within a given language or dialect. Surveys of this type seek *grammaticality judgements*, determinations of how *well-formed* sentences are, based on native speakers' knowledge of the language [30]. Input of this type is especially useful for investigating Non-Standard morphosyntactic forms which differ from more widely used Standard grammatical constructions. In addition, participants' judgements may cluster in particular patterns, but there may also be a level of individual variation that is obscured by global measurements of grammaticality. Moreover, 'naive' native speaker respondents make judgements based on *acceptability*, subject to the influence of factors such as pragmatic plausibility, rather than pure *grammaticality*, correctness of (morpho)syntactic structure [13]. To find such individual variation, researchers may wish to seek clarification through systematic follow-up questions, which can be complex to serve in current approaches to Internet-based data collection.

Our proposed system was evaluated in a use case on the grammaticality of the *Alternative Embedded Passive* (AEP) [7], which consists of a verb such as *need/want/like* followed directly by a passive participle, in contrast to *Standard Embedded Passives* (StEP), in which the passive participle is preceded by the non-finite passive auxiliary *to be*.

- *The dog needs walked* (AEP)
- *The dog needs to be walked* (StEP)

Linguistic acceptability was tested by giving respondents a binary choice over each question, where 0 stands for '*this sentence sounds strange to me*' and 1 stands for '*this sentence sounds good to me*'.

Previous linguistic studies on the AEP (without the presence of *to be*) point to *need* being the most commonly used main verb, followed by *want* and *like* [19]. Moreover, *Inanimate* subjects seem to be more acceptable with the use of *want* and *like* in the AEP than the StEP [7]. However, these findings are based on studies conducted only on the North American population using American English, and therefore may not apply to Scottish and Northern Irish speakers who use the AEP.

3   Requirements for an Informative and Dynamic System

Given the previous system, two main requirements emerged.

An Informative System Respondents should be asked a sufficient and reasonable number of relevant questions about their grammars and few irrelevant or redundant questions. Sufficient means that we ask enough questions to address our hypotheses, and reasonable means we limit the number of questions we ask to about 30, consistent with other 'dialect' surveys [16][7]. Relevance here is:

- Testing the variables by using the researcher's annotations to ask questions including all linguistic features.
- Testing the hierarchy of acceptability of different linguistic features.
- Validation of grammatical points by checking responses about grammar rather than extraneous factors such as pragmatic plausibility.
- Filtering of questions to avoid those that are known, from prior questions, not to be appropriate to the speaker.

A Dynamic System The dynamic system should serve the purposes of the informative system. In other words, the algorithms which deliver the questions to respondents ought to do so in such a way as to realise the requirements which make the system informative. This is in contrast to typical grammaticality judgement questionnaires, where the same questions are asked in every survey, not taking into account the participants' responses. In these traditional static surveys the order of the questions is predefined or entirely randomised in advance, and therefore cannot be changed as the survey is conducted. The fixed presentation of questions does not allow for a more tailored experience for the respondent and does not allow for user feedback in the form of comments to be taken into account. Additionally, the number of questions is limited, which means that a researcher may only be able to cover a select few variables of interest.

4   Knowledge Graph and Algorithms

In this section, we will present some knowledge graph based algorithms for informative and dynamic survey systems. Such survey systems are based on the notion of responsive sentence selection. In other words, the proposed survey system is able to dynamically select the next survey question, depending on the

judgement of the previous question. We will first present the two ontologies as the schema of the knowledge graph of the survey system, which allows the kind of responsive sentence selection to be presented in the algorithms in sub-section 4.2.

4.1 Knowledge Graph

Two key ontologies are designed for the proposed system: a general purpose Survey Ontology and a domain specific ontology, such as a Linguistic Feature Ontology.

The Survey Ontology contains classes such as SurveyQuestion, AnswerOption, SurveyAnswer and User, Participation, Hypothesis. It contains properties, such as hasSurveyUser, hasSurveyQuestion and hasSurveyAnswer. We refer the reader to [31] for more details of the Survey Ontology.

The Linguistic Feature Ontology has classes such as, Sentence, POS, Subject (Subject $\sqsubseteq$ POS), AnimateSubject (AnimateSubject $\sqsubseteq$ Subject), InanimateSubject (InanimateSubject $\sqsubseteq$ Subject), DefiniteSubject (DefiniteSubject $\sqsubseteq$ Subject), IndefiniteSubject (IndefiniteSubject $\sqsubseteq$ Subject), Verb (Verb $\sqsubseteq$ POS), MainVerb (with instances *need/want/like*, MainVerb $\sqsubseteq$ Verb), AEP (AEP $\sqsubseteq$ POS) and StEP (StEP $\sqsubseteq$ POS). The Linguistic Feature Ontology has properties, such as hasPOS and hasString.

When a linguistic researcher annotate survey questions (such as the one containing Sentence S1, *The dog needs walked*), a set of statements will be constructed in the knowledge graph:

- (theDog, rdf:type, DefiniteSubject), (S1, hasPOS, theDog),
- (theDog, rdf:type, AnimateSubject)
- (need, rdf:type, MainVerb), (S1, hasPOS, need),
- (walked, rdf:type, AEP), (S1, hasPOS, walked).

Based on the Linguistic Feature Ontology and the above statements, we can, e.g, classify the Sentence S1 as an instance of Sentence $\sqcap$ $\exists$hasPOS.DefiniteSubject $\sqcap$ $\exists$hasPOS.AnimateSubject $\sqcap$ $\exists$hasPOS.AEP (S1 is a Sentence that has a DefiniteSubject, an AnimateSubject and contains an AEP).

If a User U1 accepts S1, the survey system will have the following extra statements:

- (P1, rdf:type, Participation),
- (P1, hasSurveyUser, U1), (U1, rdf:type, User),
- (P1, hasSurveyQuestion, S1), (S1, rdf:type, Sentence),
- (P1, hasSurveyAnswer, *accepted*).

4.2 Algorithms

In this sub-section, we will present two algorithms that are able to responsively select sentence for the next question, with the help of sentence classification discussed in the previous sub-section (cf. the discussion of S1) .

Algorithm 1 considers the effects of linguistic features such as the choice of the main verb, namely *need, like, want*, as well as whether the subject is *Animate/Inanimate* or *Definite/Indefinite*. Along with these variables, the presence/absence of the non-finite passive auxiliary *to be* gives a total of $3*2*2*2 = 24$ possible combinations of relationships between features. These variables have been explored in previous work on this construction [7]. The linguistic researchers design 6 sentences for each of the above 24 combinations, resulting in 24*6=144 sentences, which are grouped into 12 family groups, each one of them has an AEP family of 6 sentences and a StEP family of 6 sentences.

Algorithm 1: Responsive Sentence Selection

   Input:
   g: current family group, which contains a set of ranked AEP sentences g.aep and a set of ranked StEP sentences g.step;
   x: the top percentage of sentences to be considered in the first attempt;
   Output:
   result: the set of judgement results of selected questions from g

1 $result \leftarrow nil$;
2 $s_1 \leftarrow$ random-top(g.aep,x); //randomly select a sentence from the top $x\%$ of questions as the first question
3 $result \leftarrow result \cup (s_1, judgement(s_1))$;
4 **if** *judgement(s_1) = rejected* **then**
5     $s_2 \leftarrow$ random-top(g.aep,x); //randomly select another sentence from the top $x\%$ of questions as the second question
6     $result \leftarrow result \cup (s_2, judgement(s_2))$;
7   **if** *judgement(s_2) = accepted* **then**
8        $s_3 \leftarrow$ random-bottom(g.aep,x); //randomly select a sentence from the bottom $(100 - x)\%$ of questions as the third question
9        $result \leftarrow result \cup (s_3, judgement(s_3))$;
10   **end**
11 **else**
12    $s_2 \leftarrow$ random-bottom(g.aep,x); //randomly select a sentence from the bottom $(100 - x)\%$ of questions as the second question
13    $result \leftarrow result \cup (s_2, judgement(s_2))$;
14 **end**
15 $s \leftarrow$ random-top(g.step,x);//randomly select a sentence from the top $x\%$ of questions as the final question from the StEP family
16 $result \leftarrow result \cup (s, judgement(s))$;
17 **return** $result$;

The key challenge is how to select some of the 144 sentences into a survey, which typically includes about 30 questions. The main idea is to use results form some baseline studies of these 144 sentences to learning the acceptability ranking of these sentences and related families. Instead of covering every one of the 144 sentences, Algorithm 1 selects the next sentence based on user judgements of

the current sentence (lines 4 and 7), resulting in having 2-4 sentences per family group.

Algorithm 2: Responsive Sentence Selection with Comment Understanding

Input:
g: current family group, which contains a set of ranked AEP sentences g.aep and a set of ranked StEP sentences g.step;
x: the top percentage of sentences to be considered in the first attempt;
SV: a queue of additional sentences with extra variables to be used;
SR: a queue of additional relaxed sentences to be used;
Output:
$result$: the set of judgement results of selected questions from g

1   $result \leftarrow nil$;
2   $s_1 \leftarrow$ random-top(g.aep,x); //randomly select a sentence from the top $x\%$ of questions as the first question
3   $result \leftarrow result \cup (s_1, judgement(s_1))$;
4   **if** $judgement(s_1) = rejected$ **then**
5      $s_2 \leftarrow$ random-top(g.aep,x); $result \leftarrow result \cup (s_2, judgement(s_2))$;
6   // If both s_1 and s_2 are rejected for the same reason, the next additional relaxed sentence
7   **if** $judgement(s_2) = rejected$ and $commenttype(s_1) = commenttype(s_2)$ **then**
8         $s_3 \leftarrow$ next(SR); $result \leftarrow result \cup (s_3, judgement(s_3))$;
9   **if** $commenttype(s_1) = 'to\ be'$ **then**
10   **foreach** $sentence$ s_i in $g.step$ **do**
            $result \leftarrow result \cup (s_i, accepted)$;
11   **end**
12   **if** $commenttype(s_1) = other$ and $commentMain(s_1) \neq g.mainWord$ **then**
13   **foreach** $sentence$ s_i in $similar(g, commentMain(s_1))$ **do**
            $result \leftarrow result \cup (s_i, accepted)$;
14   **end**
15   **else**
16         $s_3 \leftarrow$ random-top(g.aep,x); $result \leftarrow result \cup (s_3, judgement(s_3))$;
17   **end**
18 **else**
19      $s_2 \leftarrow$ next(SV); $s_3 \leftarrow$ next(SV); //Present the next 2 sentences with an additional variable
20      $result \leftarrow result \cup (s_2, judgement(s_2)) \cup (s_3, judgement(s_3))$;
21 **end**
22 $s \leftarrow$ random-top(g.step,x);//randomly select a sentence from the top $x\%$ of questions as the final question from the StEP family
23 $result \leftarrow result \cup (s, judgement(s))$;
24 **return** $result$;

In the setting of Algorithm 2, the linguistic researchers decide to drop the distinction of *Definite/Indefinite* Subjects, resulting in having 6 family groups, each of them has an AEP family of 12 sentences and a StEP family of 12 sentences. In order to consider even more candidate sentences (such as those in SV and SR of Algorithm 2) in surveys, Algorithm 2 allows the process of comments (lines 9 and 12) provided by users, so as to speed up the decision process: each family has 4 sentence slots; if some of these slots are not needed, additional sentences from SV and SR will be used. Consequently, Algorithm 2 allows the consideration of all the 144 sentences, as well as some additional sentences.

Note that some linguistic based optimisations and randomisation are used in the two algorithms. In order to adapt these algorithms to another domain or scenario, some per-subject optimisations and randomisation should be applied.

5   Case Studies and Evaluations

We present two case studies and associated evaluations of the algorithms.

5.1   Hypotheses

In order to address the requirements (informative and dynamic) and evaluate a tool that attempts to address them, two versions of the survey were implemented and conducted. The first version was found not to deliver sufficiently informative results, leading to development of a second version with adjustments to the algorithm employed; while the second version is an improvement over the first, we later discuss further refinements.

As described above, our case study examines the use of Alternative Embedded Passives, in which a verb such as *need, want* or *like* is followed directly by a passive participle, without the non-finite auxiliary *to be* found in Standard Embedded Passives.

– *The cat needs fed* (AEP)
– *The cat needs to be fed* (StEP)

The AEP has been claimed to be found among speakers in Scotland and Northern Ireland, but there has been little investigation of this feature for these populations. We therefore seek to investigate the following hypotheses:

– **Hypothesis 1**: Speakers who use AEP *like* will also use AEP *want*, and speakers who use AEP *want* will also use AEP *need*.
– **Hypothesis 2**: Some subset of speakers will allow inanimate subjects with AEP *want* and *like*, but not StEP *want* and *like*. Speakers who allow inanimate subjects with StEP *want* and *like* will also allow them with AEP *want* and *like*.

5.2 Case Study 1

Experiment Setup Based on the results from [31], a pool of 144 sentences were divided into 24 families, paired into 12 groups comprising both AEP and StEP sentences. The sentences in each group shared the same set of linguistic features: main verb (*need*, *want*, *like*), subject (in)animacy, and subject (in)definiteness. For instance, the group for *need*, animate subject, and definite subject included the following sentences.[4]

- *The trees need pruned*
- *The house needs painted*
- *The windows need cleaned*
- *The plant needs to be watered*
- *The garden needs to be tended*
- *This room needs to be tidied*

The sentences were ranked according to their mean ratings in the baseline results from [31], which had 50 participants over six versions, each consisting of 24 sentences covering all combinations of the main verb, (in)definiteness, (in)animacy, and [$\pm to\ be$] variables. They were presented to participants according to Algorithm 1.

The family groups were ordered to present those with main verb *need*, followed by those with main verb *want*, followed by those with main verb *like*. For each rejected sentence participants were asked 'What would you say instead?'.

Forty-six participants, who were recruited through word of mouth and social media, completed the survey online. Each answered a minimum of 24 questions; those who chose to continue could answer up to 30 questions. At the end of the survey participants were provided with an individualised map comparing their answers on one of the AEP sentences (without *to be*) with other users who had made judgements on sentences with the same set of linguistic features. See Figure 1.

Hypothesis Testing The survey system has allowed examination of Hypothesis 1 in relation to individual speakers, rather than just over global percentages. Of 46 participants, 42 accepted AEP *need*. Thirty-eight of these participants accepted AEP *want*. Ten of these 38 participants accepted AEP *like*. There were no participants who accepted AEP *like* but not AEP *want*, or who accepted AEP *want* but not AEP *need*. These results therefore confirm the hypothesis that acceptance of *want* in this construction is a precondition for acceptance of *like*, and acceptance of *need* is a precondition for acceptance of *want*.

The system also allows testing of Hypothesis 2. Fifteen participants were asked to judge sentences with inanimate subjects for *want* and *like* in both the AEP and StEP constructions; a further four were asked to judge sentences with *want* in the AEP and StEP constructions, but not *like*.

---

[4] While the sentences may vary in singular/plural subject, this is not a relevant experimental variable, but provided only for variety.
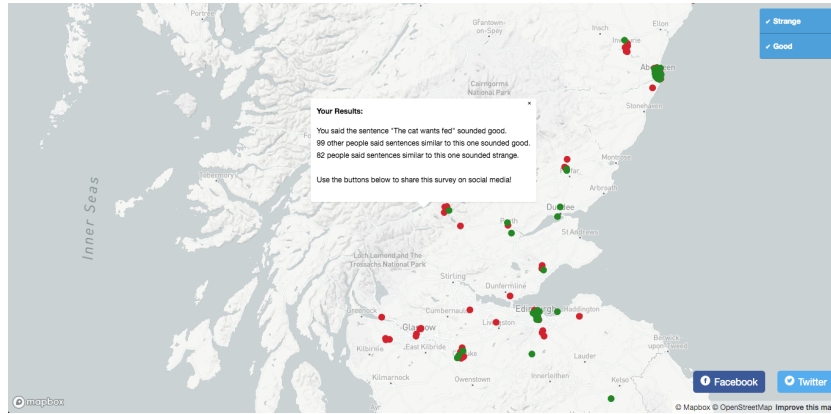
Fig. 1. Individualised Map

Of these nineteen participants, four accepted an inanimate subject with AEP *want* but not StEP *want*, while one accepted an inanimate subject with StEP *want* but not AEP *want*, contrary to the expected pattern. The rest of the speakers either accepted or rejected all inanimate subjects in both constructions. For AEP *like* two speakers accepted StEP *like* with an inanimate subject, but rejected AEP *like* with an inanimate subject, also contrary to the expected pattern. Again, the rest either accepted or rejected inanimate subjects in both types of construction.

Dynamicity and Informativity The dynamic approach used in this survey was effective in testing the *need > want > like* hierarchy of acceptance for the AEP (Hypothesis 1), in that it allowed all speakers to be asked questions for each of these main verbs, although the ordering of questions to prioritise *need* over *want* and *want* over *like* means that speakers may have been asked fewer questions overall about main verb *want* and especially *like*, as depending on their answers they may have been questioned about as few as eight family groups. For Hypothesis 1 we can therefore conclude that the survey was sufficiently informative.

At the same time, the algorithm used in this iteration of the survey, along with the limitation on the number of questions, meant that testing of Hypothesis 2 was limited. Several participants who did not use AEP *like* or *want* at all were asked to give judgements on this construction with an inanimate subject, resulting in the collection of data irrelevant to our hypothesis. Only 15 of the 31 participants who used AEP and StEP *want* were asked to give judgements on this verb with inanimate subjects; some of them also gave judgements on only a single sentence for the StEP or the AEP with an inanimate subject. The algorithm used in this iteration of the survey therefore failed to collect optimal data for testing Hypothesis 2, and was thus insufficiently informative.

The dynamic aspect of the survey was therefore partially successful. It allowed relatively strong confirmation of use of the AEP, as many speakers did not accept all sentences for this construction; had they been asked only a single sentence and rejected it the result would have been a false negative for use of the AEP. In other instances, though, the dynamic presentation of questions meant the survey collected superfluous or insufficient data.

The elicitation of superfluous judgements is a feature inherent to static surveys (i.e. those with a fixed set of questions for all respondents), and so in this respect the dynamic survey was still superior, as it eliminated these irrelevant questions for at least some participants. Unintentionally insufficient coverage of variables is a problem more easily avoided in a static survey, although by nature having a fixed set of questions circumscribes how many linguistic features a researcher can include in a questionnaire of this type. Below we will discuss amendments intended to remedy this problem in a second iteration of the dynamic survey.

5.3   Case Study 2

Experiment Setup The same set of 144 sentences was used, divided into 12 families, paired into six groups, based on main verb and (in)animacy: (in)definiteness was not used as a variable, as it was deemed irrelevant to any hypotheses of interest.

A further 18 sentences were added to the set of possible questions in order to test a number of additional variables: use of adverbs with the AEP (e.g. *The books need sorted alphabetically)*; use of *by*-phrases (e.g. *My car needs checked by a mechanic*; use of purpose-clauses (e.g. *The screws need tightened to hold the shelf up*); questions (e.g. *Does the door need opened?*); negation (e.g. *Those carpets don't need cleaned*); and relative clauses (e.g. *Those are the shirts that need ironed*). These additional linguistic features were included to measure a number of other hypotheses examined in previous work, though which are tangential to the hypotheses we address in this paper.

In this iteration the system was coded to recognise comments in response to 'What would you say instead?', in particular, the use of *to be* or an alternative main verb *need*, *want* or *like*. The sentences were presented according to Algorithm 2, again using participants' judgements from the baseline survey for ranking of the 144 original sentences.

Fifty-three participants were recruited through paid social media advertising which targeted users in Scotland and Northern Ireland. Each participant gave judgements on 24 sentences and, as in Case Study 1, was presented with an individualised map upon completion of the questionnaire and encouraged to share the survey on social media.

Hypothesis Testing Again, the system allowed testing of Hypothesis 1, that use of AEP *need* is a precondition for use of AEP *want*, and AEP *want* is a precondition for AEP *like*. Forty-six participants of 53 accepted AEP *need*, and

42 of these accepted AEP *want*. Of these 42, 17 accepted AEP *like*. A single participant appeared to accept AEP *want*, but not AEP *need*. However, closer inspection revealed that they did accept several of the sentences with additional variables (e.g. use of *by*-phrases), all of which had AEP *need*, and so did not contradict this hypothesis.

Of 35 participants who accepted both AEP and StEP *want*, 34 were asked about these with inanimate subjects. Five accepted an inanimate subject with AEP *want*, but not StEP *want*; two rejected an inanimate subject with AEP *want* but accepted one with StEP *want*. Of 15 participants who accepted both AEP and StEP *like*, 13 were asked about these with inanimate subjects. Two accepted an inanimate subject with AEP *like*, but not *StEP like*; one rejected an inanimate subject with AEP *like* but accepted one with StEP *like*. The rest of the speakers either accepted or rejected all inanimate subjects in both constructions for *want* and *like*. These results therefore weakly support Hypothesis 2, that inanimate subjects are more acceptable for *want* and *like* in AEP constructions than StEP constructions.

Dynamicity and Informativity Because the sentences were divided into fewer families, and participants were not questioned about lower-ranked sentences, this survey was more effective in testing both hypotheses. The addition of comment understanding also meant that some questions could be eliminated, as participants' acceptance of StEP forms (with *to be*) and alternative main verbs could be confirmed by their responses to 'What would you say instead?'. As a result, it was possible to include additional sentences, testing more variables for many respondents.

While this iteration of the dynamic survey therefore addressed the problem of insufficient coverage of variables, thereby increasing informativity, presentation of superfluous questions to some participants still remained a problem. In particular, consistently 'Standard' speakers, i.e., those who do not use the AEP at all, were repeatedly presented with AEP sentences because their input of *to be* forms for 'What would you say instead?' meant that these StEP forms were marked as grammatical, triggering the algorithm to bypass [3]. These speakers were therefore never presented with StEP sentences. Instead, they were asked questions about the sentences with additional variables, all of which were AEP forms.

This issue was highlighted by a response left on social media by a participant that the questionnaire became 'boring' because all of the questions seemed to require 'the same grammatical addition'. Decreasing participant interest as the survey progresses therefore remains a problem, although it is notable that some 'Standard' speakers did complete the survey; this problem is potentially remedied by reducing the overall number of questions for participants whose answers indicate that they do not use the Non-Standard form, or introducing new variables and/or constructions for such users in order to increase participant engagement and level of informativity for researchers.

6 Related Work

6.1 Intelligent Surveys

One of the intelligent surveys systems already implemented is the *Dynamic Intelligent Survey Engine* DISE [29], which aims to have an as flexible as possible approach to creating a survey while avoiding being restricted. Similarly to our old system, it uses a wide variety of data methods and an advanced data collection approach with the intent to measure the consumer preferences. However, in contrast to our system, which uses a drag and drop interface for creating surveys, survey creating in their system is done by XML markup language, which may have a rather steep learning curve and thus cumbersome to learn. Furthermore, the system does not allow for conditional trigger for better user experience, nor does it use its knowledge to prioritise the most significant questions first.

6.2 Psycholinguistic Surveys

MiniJudge [22] is a tool specifically designed for theoretical syntacticians to help them design, run and analyse judgemental experiments in the minimum amount of time with maximised efficiency and without any prior training. This is achieved by 'minimalist' experimental syntax where experiments are conducted on a small participant group, sets of questions and quick a survey. Moreover, it offers automation of the statistical analysis of data, and thus is beginner friendly.

WebExp [17] is a software package to run psychological experiments over the internet and measure the respondents reaction time (latency). The system shows a nuance approach on collecting latency measurements and replicating lab-based conditions accurately across multiple platforms. Similarly *PsyToolkit*[5] has been specifically designed to setup, run and analyse questionnaires and reaction time experiments. Furthermore, the system links the experiments online, which can be easily embedded in social media networks and used for participant recruitment [33].

Other psycholinguistic tools include IBEX [6] *Internet Based EXperiments* which focuses on grammatical judgements. The questionnaire is presenting the sentences in a different variety of ways; *FlashSentence* method where sentences are 'flashed' to the participant for only a limited amount of time and *Dashed-Sentence* method where the sentences are presented either chunk-by-chunk or word-by-word. On the other hand, in order to make the research feasible across different groups with individual languages *Wordlikeness* allows to design questionnaires with text, audio, images and video files [4].

Lastly, many other popular survey system are used for information gathering in a field of linguists. One of which is *SurveyMonkey*[6] which allows people to develop surveys online, deploy it and test it to the community and then analyse collected data. Another one being *Amazon's Mechanical Turk (MTurk)*[7], where

---

[5] https://www.psytoolkit.org/

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

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

people complete surveys for money. Johnson suggests that using such a platform can provide a large participation-pool with necessary tools to build an experiment in quick and efficient manner [14]. *Turkolizer* [11] and *Turktools* [9] are two tools that run on this crowd sourcing platform. While this approach may potentially present benefits in large-scale experiments, this platform presents only a basic statistical analysis of the data. To do any form of knowledge powered services, for instance syntactic and semantic evaluation of the results, a knowledge structure would have to be implicitly hard-coded. As a result of this, the experiments data is hard to transmit, link or reuse.

6.3   Commercial Surveys

Whilst not as related to other intelligent / linguistic surveys, commercial surveys can provide a great insight on other aspects of surveys such as: *user interfaces, security* and *distribution* [3,32]. Many of which (aspects) are highly paramount as their primary goal is to attract as many customers as possible.

6.4   Adaptive Questionnaires

The use of adaptive questionnaires is a widely used concept in identifying learning styles of students. The most popular approach of learning style recognition is via the use of questionnaires. While they might be effective, they have disadvantages: (1) filling a questionnaire is time-consuming since questionnaires usually contain numerous questions; (2) learners may lack time and motivation to fill in long questionnaires; and (3) a specialist needs to analyse the answers. [1]

Several questionnaire systems have been proposed to mitigate the above stated issues and automatically minimise the number of questions using various algorithms. *AH questionnaire* [24] used decision trees as the main algorithm and managed to reduce the number of questions by over 50% and achieved over 95% accuracy when predicting the students learning preference. A tool proposed by Noke-lainenet al. [23] uses Bayesian modelling as well as abductive reasoning and accomplished similar question reduction of 50% as in previous system. More relevant work to our system has been done with *Q-SELECT* [20] using neural network and decision trees to decrease the number of questions by trying to find the least influential question in the survey. Furthermore, it is capable of reordering the questions and thus provide a personalised questionnaire to the end-user. Recently, their system *T-PREDICT* has been further improved from 35% reduction of questions to over 85% reduction, while keeping the error rate comparable at only 12.1% [21]. However, none of these approaches address how to responsively select the next question given the judgement of the current question.

7   Conclusion

With the help of Knowledge Graph, we propose a dynamic approach to the questionnaire component of the survey, yielding more informative results. The

questions are ordered by a model based on their importance. Once the questions are ordered, a set of conditional triggers are set to provide a more dynamic experience, which benefits the researcher in maximising the quality and quantity of data collected, and the user in creating a more varied survey. Follow-up questions are asked in a case of the user accepting or rejecting certain questions.

In the evaluation we have shown that the dynamic component can have a positive impact on the quality of the data as well as limiting the number of questions asked in the survey. The previous system performed 6 different surveys, each of which had 24 questions; a total of 50 people participated in those surveys. With our system we have managed to achieve the same results as the previous study in the one iteration of the survey, asking 28.2 questions on average with the same types of questions having only 25 participated in the survey. Such improvement is based on the semantic understanding of survey questions enabled by knowledge graphs.

References

1. Abernethy, J., Evgeniou, T., Vert, J.P.: An optimization framework for adaptive questionnaire design. INSEAD, Fontainebleau, France (2004)
2. Callegaro, M., Wells, T., Kruse, Y.: Effects of precoding response options for five point satisfaction scales in web surveys. In: 2008 PAPOR Conference. Citeseer (2008)
3. Capterra: Survey software buyers' guide. https://www.capterra.com/survey-software/#buyers-guide (2019), [Online; Accessed: 5 Mar, 2019]
4. Chen, T.Y., Myers, J.: Worldlikeness: a web-based tool for typological psycholinguistic research. University of Pennsylvania Working Papers in Linguistics **23**(1), 4 (2017)
5. Dolnicar, S., Grün, B., Yanamandram, V.: Dynamic, interactive survey questions can increase survey data quality. Journal of Travel & Tourism Marketing **30**(7), 690–699 (2013)
6. Drummond, A.: Ibex 0.3. 7 manual (2013)
7. Edelstein, E.: This syntax needs studied. Micro-syntactic variation in North American English pp. 242–268 (2014)
8. Elmes, D.G., Kantowitz, B.H., Roediger III, H.L.: Research methods in psychology. Cengage Learning (2011)
9. Erlewine, M.Y., Kotek, H.: A streamlined approach to online linguistic surveys. Natural Language & Linguistic Theory **34**(2), 481–495 (2016)
10. Evans, J.R., Mathur, A.: The value of online surveys. Internet research **15**(2), 195–219 (2005)
11. 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)
12. Guin, T.D.L., Baker, R., Mechling, J., Ruyle, E.: Myths and realities of respondent engagement in online surveys. International Journal of Market Research **54**(5), 613–633 (2012)
13. Here, M., Now, P.: Bing, bang, bong. Blah **00**(0), 00–00 (0000)
14. Johnson, D.R., Borden, L.A.: Participants at your fingertips: Using amazons mechanical turk to increase student–faculty collaborative research. Teaching of Psychology **39**(4), 245–251 (2012)

15. Kaminska, O., McCutcheon, A.L., Billiet, J.: Satisficing among reluctant respondents in a cross-national context. Public Opinion Quarterly **74**(5), 956–984 (2010)
16. Katz, J.: The british-irish dialect quiz. New York Times **00**(0), 00–00 (15 February 2019)
17. 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)
18. Kropf, M.E., Blair, J.: Eliciting survey cooperation: Incentives, self-interest, and norms of cooperation. Evaluation review **29**(6), 559–575 (2005)
19. Murray, T.E., 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)
20. Mwamikazi, E., Fournier-Viger, P., Moghrabi, C., Barhoumi, A., Baudouin, R.: An adaptive questionnaire for automatic identification of learning styles. In: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. pp. 399–409. Springer (2014)
21. Mwamikazi, E., Fournier-Viger, P., Moghrabi, C., Baudouin, R.: A dynamic questionnaire to further reduce questions in learning style assessment. In: IFIP International Conference on Artificial Intelligence Applications and Innovations. pp. 224–235. Springer (2014)
22. Myers, J.: Minijudge: Software for small-scale experimental syntax. International Journal of Computational Linguistics & Chinese Language Processing, Volume 12, Number 2, June 2007 **12**(2), 175–194 (2007)
23. Nokelainen, P., Niemivirta, M., Kurhila, J., Miettinen, M., Silander, T., Tirri, H.: Implementation of an adaptive questionnaire. In: Proceedings of the ED-MEDIA Conference. pp. 1412–1413 (2001)
24. Ortigosa, A., Paredes, P., Rodriguez, P.: Ah-questionnaire: An adaptive hierarchical questionnaire for learning styles. Computers & Education **54**(4), 999–1005 (2010)
25. Pan, J., Calvanese, D., Eiter, T., Horrocks, I., Kifer, M., Lin, F., Zhao, Y.: Reasoning Web: Logical Foundation of Knowledge Graph Construction and Querying Answering. Springer (2017)
26. Pan, J., Vetere, G., Gomez-Perez, J., Wu, H.: Exploiting Linked Data and Knowledge Graphs for Large Organisations. Springer (2016)
27. Puleston, J., Sleep, D.: The game experiments: researching how gaming techniques can be used to improve the quality of feedback from online research. In: Proceedings of ESOMAR Congress (2011)
28. Saleh, A., Bista, K.: Examining factors impacting online survey response rates in educational research: Perceptions of graduate students. Journal of MultiDisciplinary Evaluation **13**(29), 63–74 (2017)
29. Schlereth, C., Skiera, B.: Dise: dynamic intelligent survey engine. In: Quantitative marketing and marketing management, pp. 225–243. Springer (2012)
30. Schütze, C.T.: The empirical base of linguistics: Grammaticality judgments and linguistic methodology. Language Science Press (2016)
31. Soares, R., Edelstein, E., Pan, J.Z., Wyner, A.: Knowledge driven intelligent survey systems for linguists. In: Joint International Semantic Technology Conference. pp. 3–18. Springer (2018)
32. SoftwareAdvice: Buyer's guide. https://www.softwareadvice.com/za/survey/#buyers-guide (Mar 2019), [Online; Accessed: 23 April, 2019]
33. Stoet, G.: Psytoolkit: A novel web-based method for running online questionnaires and reaction-time experiments. Teaching of Psychology **44**(1), 24–31 (2017)