Is a Hot Dog a Sandwich?: Generating Conversation Starters That Categorize a Word in an Unusual Way

KATJA WEVERS, University of Twente, The Netherlands

Conversations are a part of everyday life, but sometimes finding a way to start a conversation can be difficult. Having a fun question that can act as an ice breaker can not only help humans to start a conversation, but can also be used for virtual agents to simulate a natural interaction. We propose a system that can generate a creative question that can spark a discussion by trying to put a word into an unusual category. Evaluation shows this system is able to create such a conversation starter, but not *all* output it creates is successful. The program can still be improved in creativity and efficiency.

Additional Key Words and Phrases: computational creativity, conversation starter, ontology

1 INTRODUCTION

Creativity is often seen as an inherent part of human intelligence [4], but researchers have been trying to investigate whether it can be broken down in a structural way, creating the potential for creative agents to be computationally programmed. This field of research, labeled computational creativity, is a field that has only arisen in recent years, but is rapidly growing. Computational creativity can be applied in many creative domains, such as music, story, language, and humor [3]. One specific area within the language domain of computational creativity that is not as well-researched is the area of conversation topics. In fact, neural generation models for dialogue in general are still not fully understood [5]. To keep the scope maintainable, we stay within the specific topic of starting conversations.

Whether it is at the coffee machine or on the street, conversations are a part of daily life, but sometimes it can be hard to get a conversation going. Especially when meeting new people, it can be hard to think of a good conversation topic. One way to start a conversation is by asking a question. Research has shown that question-asking can be a good way of keeping a conversation [5], but how does one think of a good question? This is where a computer program could potentially help. Making a program that can generate questions could provide a stepping stone to start a conversation. An important part of conversation is having a back and forth in who is talking [5]. Therefore, the question that is to-be-generated should be able to spark a discussion. *Is a hot dog a sandwich?* is, though somewhat silly, one such example. While for some people the immediate answer is no, arguments could be made on why it can be classified as a sandwich. The question uses one of the qualities of a hot dog (e.g., having bread) to try and classify it into a certain category. Having a set of questions that can spark such a discussion, especially a humorous one, can be a nice way to start a conversation or simply break the ice.

TScIT 37, July 8, 2022, Enschede, The Netherlands

In this paper we present a computer system that generates questions similar to the previously mentioned one. These questions try to categorize words in an unusual way, with the goal for these questions to be interesting enough to spark a discussion. The categorization is based on the aspects and semantic relations of nouns. First, we explore the field of computational creativity more to see what similar work has been done that can inspire this paper's work. We then explain the problem in more detail so a method of creating the system can be set up. Finally, the generated questions are evaluated to see how well they would do at starting a conversation.

2 PROBLEM STATEMENT

This research aims to combine the field of computational creativity with the topic of starting conversations. To keep the scope at a reasonable level for the given time frame, we zoom in on a specific way to start a conversation: a fun question. We then focus on one specific type of question, asking about an object, or more generally, a noun, trying to classify it in a new or unusual way. To solve this problem, we ask the following Main Research Question (MRQ):

- **MRQ**: How do we define and implement a computer system that can generate conversation starters that categorize a word in an unusual way?

To answer this question we break the process down into smaller parts, resulting in the following three Sub Research Questions (SRQ):

- **SRQ1**: What structure can be identified to create the conversation starter?
- **SRQ2**: What are good characteristics of a conversation starter?
- **SRQ3**: To what extent is the output of the computer program able to start a conversation?

3 RELATED WORK

This research mostly uses the research of other domains than the one of conversation starters as inspiration to solve the problem. There are multiple domains within computational creativity that focus on creating some sort of textual output. The most relevant is the domain of jokes and humor, since the nature of the proposed question generation can be seen as humorous. Different systems and structures have been created that can create a humorous textual output. Witscript is a system that generates improvised jokes, with the goal to be able to use those jokes to make chatbots more human-like and likeable [8]. The jokes are based on the context of the chat conversation that is being held, using word associations to create a punch line. In a similar line, we look at a Pokémon generator, which generates a new Pokémon by blending words provided through user input and makes a description using those words and their semantic relations [2]. Both of these systems look at word definitions, associations and semantic relations to create their output.
A general humor generation model has been created which presents a generator that has a dynamic model and dynamic parameterization,

allowing improvement through learning from example jokes [10]. While a model with non-fixed elements like this is more promising in terms of creativity, it falls outside the scope of this research.

In addition to these generators, we examine the generation of topics and questions. Sevegnani et al. propose a system that can generate questions based on a piece of text about a certain topic [1]. Next to this, a system was made that could generate a piece of text that can transition from one topic to another [6]. However, both of these papers are made with the goal of improving conversations between humans and chat bots, whereas the focus of this paper lies on real-life conversations.

Lastly, we consider a paper that aids in the conversational aspect of this research. Research has been done to investigate what constitutes as a conversation [5]. It considers attributes of conversation, as well as how humans evaluate the quality of a conversation. The attributes considered most important to a conversation are repetition, specificity, response-relatedness, and question-asking. The last option is what our system will help stimulate. Humans evaluate the quality of a conversation on multiple aspects, where interestingness and making sense are most relevant to this research, as other aspect include ones such as listening and inquisitiveness, which are unattainable for our system.

4 METHODOLOGY

To create the program, a structure first needs to be identified that can define the nature of the conversation starter. This structure can then be translated into a Python algorithm which should be able to generate a conversation starter as an output. Finally, a survey is held to evaluate the overall quality of the generated conversation starters.

4.1 Generating conversation starters

4.1.1 Structure. We create the structure by taking the example from the introduction, *Is a hot dog a sandwich?*, and looking at the qualities of both the question itself and the two concepts, *hot dog* and *sandwich*, it contains. The question tries to place the instance *hot dog* in the more general category of *sandwiches*. Thus, we can say that one of the two concepts in question should have different types, thereby forming a category that the instance could fit into. Additionally, we find that the reason why the question is good at starting a conversation is because opinions are generally divided about its answer; there are arguments both for and against it. Therefore, there should be similarities between the concepts, but not enough for the generated question to have an obvious answer. This information can be summarized into the following list of aspects for two words to be considered good concepts for the structure:

- One of the two concepts should be general enough to be considered a category
- The two concepts share some common features, in particular the instance shares some (but not all) properties with all or most instances belonging to the category
- The instance is not already commonly considered to be a part of the category

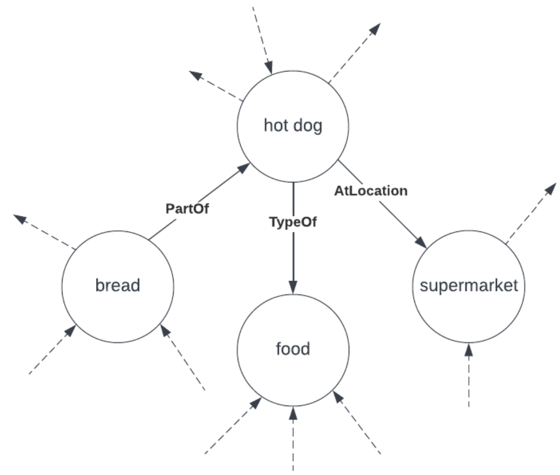

Fig. 1. Example of ConceptNet graph for the word *hot dog*. The amount of edges has been greatly reduced for simplicity.

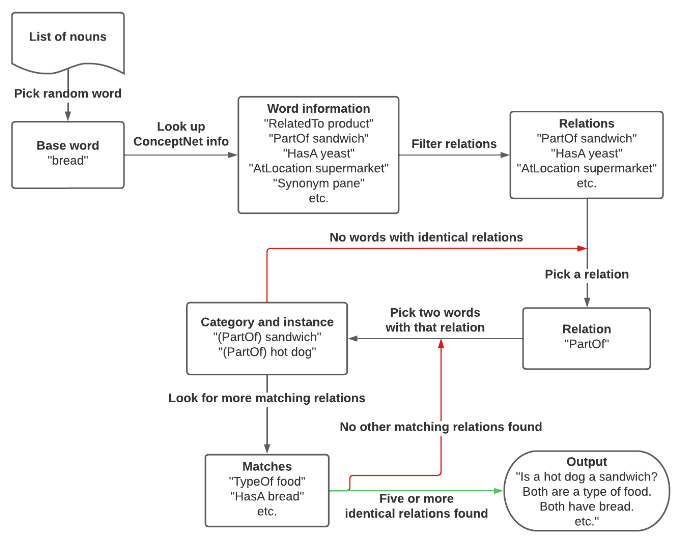

Fig. 2. Diagram showing the process of finding conversation starters

4.1.2 Algorithm. With this list of criteria the algorithm can be created. The first element that is essential is having information about the aspects and semantic relations of words.

Retrieving word information Multiple resources exist that can provide a database of words and their definitions and semantic relations to other words. After comparison and close consideration, ConceptNet[1] was chosen for this project due to its open-source

[1]ConcepNet can be found at https://conceptnet.io/. ConceptNet is a semantic network of nodes and edges where each node represents a word and each edge a relation. It has many types of relations, such as the common related terms, its types, but also things like the locations of where the word can be found. For example, looking up *hot dog*, will show that it can be found at a fair or at a fast-food restaurant [7].

availability and how extensive its knowledge on words and their relations is. A visualization of a small part of ConceptNet's information about the word *hot dog* can be seen in Figure 1 as an example.

We explain the algorithm using Figure 2, which shows how the algorithm works using the example of how the question *Is a hot dog a sandwich?* could potentially be generated.

Start The program starts with a list of common nouns[2] from which one random word is picked. In this case, the word *bread* is picked. This word forms the base node where we will explore outwards from. The algorithm is designed in such a way that the word that is picked forms the first common relation between the concepts that are used for the final generated conversation starter. In our example, this common relation is *PartOf*, since *bread* is both a *PartOf* a *hot dog* and a *PartOf* a *sandwich*.

Filtering word information Relations like these are gathered using ConceptNet. Data from ConceptNet is crowd-sourced, so is not always reliable. ConceptNet gives a weight to each relation to show how believable the information is. It correlates to the amount of sources that have been collected that support the information, as well as the reliability of the sources. To reduce the amount of unreliable information, edges with a weight less than 1.0 are filtered out. Additionally, not all relation types that ConceptNet offers are relevant for our research, such as the *RelatedTo* relation, which holds all general words that are related but for which it is not known what kind they are exactly. These relations are filtered out as well.

Selecting an instance and a category Now that only the information of the word that is useful to us is left, we can explore its edges. The algorithm look at all the relations that it has for *bread* and picks a random one, in this example *PartOf*, for which it will pick the first two words it finds that both have that relation.

Comparing relations In the best-case scenario, it picks *hot dog* and *sandwich*. It will then look up ConceptNet's information on both of those words, applying the same filtering as previously. The two lists of relations for these words are then cross-referenced to look for more identical relations. Identical relations for *hot dog* and *sandwich* could for example be *HasA filling* or *TypeOf food*.

If enough matches are found, the two words are considered good concepts for the conversation starter. The number five was chosen as a minimum number of matching relations required. This was done through visual inspection of the output. This is slightly in contrast with the list provided in Section 4.1.1, which states there should be a maximum amount as well to prevent too similar words to be picked. However, this was left out since it was predicted to be too time-consuming to spend time on finding an adequate maximum value.

Output Once enough matches have been found, the output can be created. The conversation starter always has the same template: *Is [word1] [word2]?* thus we get our question *Is a hot dog a sandwich?* The algorithm also formats all found identical relations into a coherent output, forming an explanation which elaborates on why it thinks the instance could be placed into the category. Further information on how the explanation is created can be found in Appendix B.1.

The algorithm does not stop after finding its first match, but rather continues exploring the next two words of the initial relation (*PartOf*), repeating the process. After all words with that relation have been explored, it will do the same for the other relations that *bread* has. Once it has gone through all those relations, all found matches are collected and a list of all found possible conversation starters are given as output, including their explanations.

[2] The nouns list was retrieved from http://www.desiquintans.com/nounlist and modified by removing unfitting words. A more detailed explanation of what words were removed can be found in Appendix A.1.

4.2 Evaluating the conversation starters

The main goal of the generated questions is that they should be able to spark a fun discussion. Once the system was refined to the best of our ability, we tested how effective it is. Ideally this would have been done using an ecological evaluation set in a real environment where the questions are used in a real-life situation. However, due to time constraints, a survey was done instead.

4.2.1 Survey. Because the focus of this research lies in the general possibility of creating the described system and not necessarily making a perfect system, the survey focused on quality instead of quantity. A set of conversation starters was chosen out of all of the generated ones, which were used in the survey. The set consists of some that are predicted by the researcher to have mixed answers, as well as some that seem to have a more obvious answer to keep a somewhat even balance. The chosen conversation starters can be found in Table 1 and their generated explanations can be found in Appendix B. An elaboration on how the survey conversation starters were selected can be found in Appendix A.2.

Table 1. Survey conversation starters

Set and question number	Conversation starter
Q1.1	Is a band an orchestra?
Q1.2	Is peanut butter a soup?
Q1.3	Is space a shape?
Q2.1	Is a town a shopping mall?
Q2.2	Is the ocean a pool?
Q2.3	Is a drawer a box?
Q3.1	Is a clarinet a flute?
Q3.2	Is a desk the office?
Q3.3	Is a pocket a purse?

For every conversation starter, participants were asked to answer the question with yes or no and to explain their answer. This would give an indication as to how the opinions are divided about the question. If the participant answered no, they were then shown the generated explanation to the question, after which they were asked again to answer the question with yes or no. This could give an indication as to whether it was possible to change people's opinion based on the explanation, giving more insight to the possibility of having a discussion.

In addition to these questions, the participants were asked to rate the question using a Likert scale ranging from 1 to 5 on how well the participant thinks they could hold a discussion about the question. They were also asked to rate the conversation starter on creativity. These last two questions create a more direct way of seeing how well the generated questions would do at starting a conversation, as they ask about the participant's opinion. Asking about the creativity aimed to see how fun the possible discussion the participant would think to be.

4.2.2 Method of evaluation. The results of the survey will be evaluated in way that is similar as done by Wiedmaier and Lardner [9], by comparing the ratio of people agreeing and disagreeing with the categorization proposed by the conversation starters. Intuitively, we can assume that the closer the division, the higher the possibility to have a discussion about it, since there is a big chance that when using it in a real-life setting, the opinions of the people present will be divided as well. The closer the division is to being 50% yes and 50% no, the more likely the generated question is to be a good conversation starter.

To support the findings from evaluating this question, we look at the explanations that participants give to their answers. Here we can see how well they can support their answer, giving an indication as to what the start of the discussion on their side would look like. Of course, this does not include any actual back and forth discussion, which is why the generated explanation is added when the participant answers no. With this, we simulate what another person could potentially say if it was a real-life discussion. This method of evaluation uses prediction of how a discussion would go without the participant necessarily being aware of the goal of asking the question.

We also use the more direct method of asking for their opinion on how well they think they could hold a discussion about it and how creative they think the generated question is. The former gives an indication of whether the discussion would be capable of back and forth and whether it will last for a while, and the latter helps indicate how fun the conversation would potentially be. With this, we can combine both our prediction and the participant's prediction to see how well the generated question might do at sparking a discussion.

5 RESULTS AND DISCUSSION

We first evaluate the survey to approximate how well the generated questions would do at actually starting a conversation. Granted, a real evaluation is still needed that looks at the output in the context of a conversation in a live setting, should we want to accurately judge the questions' ability to start a conversation. For this research, we let the survey be our rule of judgement to assess the algorithm.

5.1 Survey

The survey collected a total of 41 individual responses, where some participants filled in multiple sets of questions. Set 1 was filled in 24 times, and set 2 and 3 both 19 times.

Overall, the answers given to the conversation starters were not always clearly divided. The ratio of agreement versus disagreement for each conversation starter can be seen in Figure 3. It shows that for three of the conversation starters (Q1.2, Q2.1, and Q3.2) less than

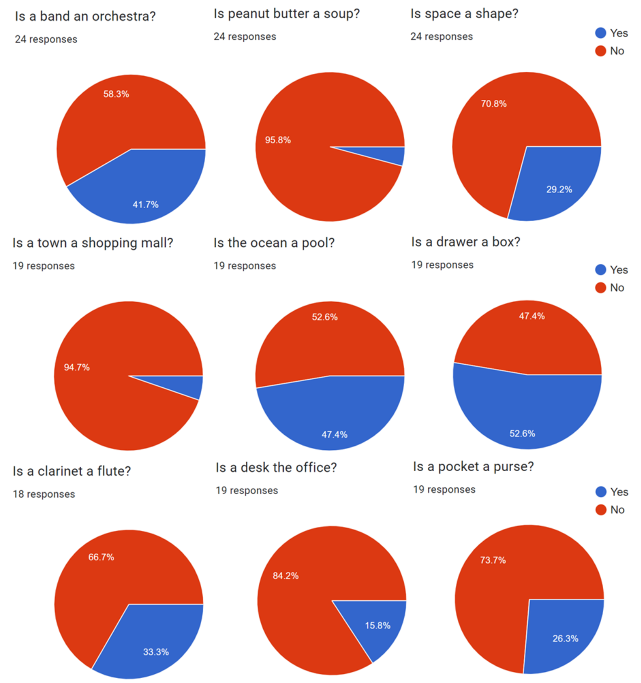

Fig. 3. Survey results for each conversation starter.

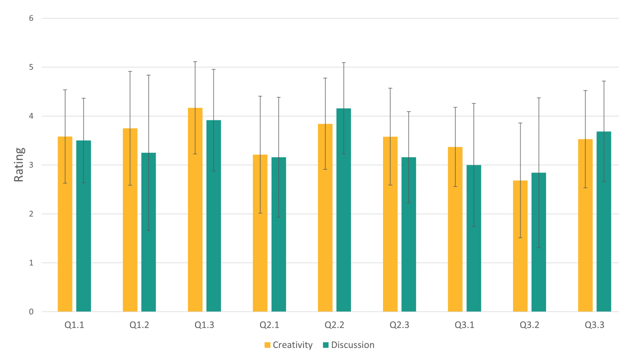

Fig. 4. Diagram showing the mean of the results from each survey question about rating the creativity and how well a discussion could be held. The black lines for each bar represent the standard deviation.

a quarter of the participants had a differing opinion from the others. Overall, this shows that the majority of the conversation starters have a divided opinion, alluding to good discussion possibilities. The open answers differed per length, but often participants tried to define both words and then point out the obvious differences. Participants who agreed tended to not be entirely convinced necessarily, but found they could come up with arguments in favor of the question and were not opposed to agreeing. We see this as one participant wrote the following response for Q3.3: "My first reaction was no, because a pocket is something integrated to your clothing, while a purse is a separate object. However, I can leave my jacket including its pockets at home, the same way I can leave a purse at home. So the purse being a pocket on its own can be replicated by any other pocket."

When looking at how many participants change their answer from disagreeing to agreeing after seeing the explanation generated by the algorithm, we find that the explanation does not prove to be helpful at all. For Q1.1, Q2.1, Q3.2, and Q3.3 one person changed their answer to agree and for Q3.1 two people changed to agree, while for all other questions, no participant changed their mind. Converting this into percentages, it shows that this equates to being less than 15% percent for all cases. This can have multiple causes. One is seen in the fact that some questions have an obvious answer if the participant has enough knowledge about the words. We see this most prominently in Q3.1. It takes certain musical knowledge to know that a clarinet uses a reed whereas a flute does not. This fact was important enough for participants answering no for the two words to be distinct.
Another cause could be the quality of the explanation. Since ConceptNet is crowd-sourced, relations exist that could be considered unusual. Not only are they not always correct (*Both are a municipality.* for Q2.1), sometimes they are quite peculiar (*Both are used for impressing a date.* for Q3.1). Explanations like these do not help in convincing someone to agree with the categorization.

Lastly, we look at the response to rating the creativity of the conversation starters and rating the ability to hold a discussion about the categorization. The mean of the results to each question can be seen in Figure 4. All conversation starters scored somewhere in the top half of the Likert scale, with the exception of Q3.2: *Is a desk the office?*, which scored slightly below the mid-point of the scale for both aspects. While the mean shows overall positive results, we can see from the standard deviation also shown in the graph that opinions are generally quite divergent. Most standard deviations are around 1, with those for the discussion rating for Q1.2 and Q3.2 actually being around 1.5. Though these deviations are on the higher side, the results still show potential for having a fun discussion about the conversation starter.

5.2 Algorithm

From the survey we find that the system is capable of creating fun questions that can start a conversation, but that not every conversation starter is as successful as intended. There are multiple points in the algorithm where this shortcoming might be caused.

5.2.1 Categorization. The main problem found in some of the generated questions is that the word that is being categorized is either very similar or already the same as the categorizing word. For example, one generated question is *Is a kitchen drawer a drawer?*. In this case it tries put a word into a category it is already in. This is partly caused by shortcomings of ConceptNet. While it is extensive, it does not have all knowledge, and in this case, a kitchen drawer is not registered as a *TypeOf* drawer. It is also partially caused by the algorithm, as it does not check for existing relations between the category and instance. This was listed as a criterion in the original list made for the structure in Section 4.1.1 but was left out due to time constraints. It was considered the least important bullet point, as it mostly causes clutter for the output by generating ineffective conversation starters which then have to be filtered out by a human.

5.2.2 Starting nouns. The algorithm is not always successful in finding any conversation starters. It was run more than 40 times to find the final survey options, often not generating any output. This happens when the starting noun is unfitting for the structure. Even though the nouns list was filtered for rude and uncommon words, these were not the only type of words that had to be filtered out. For more vague words like *validity* or *prosperity* it cannot find anything. It appears that the starting noun needs to have a more tangible description, either in a physical or mental sense. Since the nouns list contains more than 6000 words, it was chosen to run the algorithm again if nothing was found instead of filtering the list beforehand to remove such words.

5.2.3 Relation types. The algorithm looks for matching relations, but does not check whether the relation types it finds are different from each other. This results in some explanations only consisting of the same type of relation. Some relations are more general than others. Take for example the *AtLocation* relation type: if two objects can both be found at the same location, it does not necessarily mean they can be compared. Therefore, if the only matching relations found are of type *AtLocation* the instance is not valid for the category. Not taking this into account in the algorithm gave the resulting conversation starter of *Is a wiener dog a lizard?* since both can be found at a yard, at someone's house, and at an animal shelter, to name a few. This could be prevented by adding a check to the algorithm that either checks the matching relations for their diversity or to check for the appearance of at least one important relation. In the latter case, the relations would have to be further considered to see what constitutes as an important relation.

6 CONCLUSIONS AND FUTURE WORK

This study examined how to develop an algorithm that can generate a fun conversation starter that tries to categorize a word in an unusual way. Evaluation shows that the resulting system is able to produce output that can start a conversation, but it is not always successful. Human input is still necessary to filter out bad questions. From the evaluated conversation starters, 60% had a significant split in opinion. This can be linked to these questions forming good discussion topics, as they can be argued from different sides. Furthermore, all questions scored decently on creativity and the participants' ability to argue their side, showing potential for a fun conversation. Granted, not all output is eligible to start a conversation, but improving the algorithm could create the possibility for more promising output.
Improvements to the system can be made by improving its efficiency. This could be achieved by filtering the list of initial nouns or adapting the algorithm so it checks for an existing relation between the instance and the category. The algorithm can also be improved by adding a check for the existence of an *IsA* relation to ensure the instance and category are comparable in type.
One last option for improvement is to try to make generated questions more creative. A possible way to make them more fun could be

to adapt the algorithm in a way that it will look multiple steps further from the base node. Since it only looks at the relations directly related to itself, the resulting instance and category are often quite closely related, like *band* and *orchestra* or *ocean* and *pool*. While these questions can still be fun to discuss, they are relatively not very creative. Extending the algorithm so it looks deeper into the node network could provide more unexpected, creative results. However, this does decrease the efficiency, as the amount of edges the system would have to explore would increase greatly.

REFERENCES

[1] Yllias Chali and Sadid A Hasan. 2015. Towards Topic-to-Question Generation. (2015). https://doi.org/10.1162/COLI

[2] Dominique Geissler, Elisa Nguyen, Daphne Theodorakopoulos, and Lorenzo Gatti. 2020. *Pokérator - Unveil your inner Pokémon*. ACC = Association for Computational Creativity.

[3] Roisin Loughran. 2017. *Application Domains Considered in Computational Creativity*. Technical Report. https://www.researchgate.net/publication/317598403

[4] Joel Parthemore. 2021. *Will the real artist stand up? Computational creativity as mirror to the human soul*. Technical Report.

[5] Abigail See, Stephen Roller, Douwe Kiela, and Jason Weston. 2019. What makes a good conversation? How controllable attributes affect human judgments. (2 2019). http://arxiv.org/abs/1902.08654

[6] Karin Sevegnani, David M Howcroft, Ioannis Konstas, and Verena Rieser. 2021. *OTTers: One-turn Topic Transitions for Open-Domain Dialogue*. Technical Report. 2492–2504 pages.

[7] Robyn Speer, Joshua Chin, and Catherine Havasi. 2016. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. (12 2016). http://arxiv.org/abs/1612.03975

[8] J Toplyn. 2021. *Witscript: A System for Generating Improvised Jokes in a Conversation*. Association for Computational Creativity = ACC.

[9] Ben Wiedmaier and Matt Lardner. 2018. Why can't we settle the "is a hot dog a sandwich?" debate? https://dscout.com/people-nerds/is-a-hot-dog-a-sandwich-great-debate

[10] Thomas Winters, Vincent Nys, and Danny De Schreye. 2019. *Towards a General Framework for Humor Generation from Rated Examples*. Technical Report.

A FILTERING

At multiple points in the implementation process filtering of certain information is done. Since not all filtering processes were considered important to the main explanation of the methodology, some are explained below.

A.1 Filtering the initial nouns list

The nouns list has been filtered by taking out words deemed unfitting for the conversation starters. The conversation starters are intended to be fun and nonsensical, so words that have generally negative associations have been taken out, such as *depression* and *anxiety*. Furthermore, uncommon and incredibly specific words have been left out, such as *planula* and *vanadyl*. This last step was done roughly by skimming over the words.

A.2 Choosing the survey conversation starters

From the survey questions Q2.1, Q3.1, and Q3.2 were predicted to have more obvious answers, since they include words with more strictly defined descriptions (clarinet and flute) or there are clear differences between the two words which are too distinct for the two words to still be comparable (an office is a room and a desk is an object, a town has houses that people can live in and a shopping mall does not).

B ALGORITHM RESULTS

The following section shows some output generated by the algorithm, including those used in the survey.

Is a band an orchestra?
A musical instrument can be found at both a band and an orchestra.
A clarinet can be found at both a band and an orchestra.
A bass clarinet can be found at both a band and an orchestra.
A keyboard instrument can be found at both a band and an orchestra.
A trumpet can be found at both a band and an orchestra.
A piccolo can be found at both a band and an orchestra.
Both are a collar.
A trombone can be found at both a band and an orchestra.
Both are a part of section.
A tuba can be found at both a band and an orchestra.
Both are a musical organization.
A cornet can be found at both a band and an orchestra.
Both are a restraint.

Is peanut butter a soup?
Both can be found at a jar.
Both can be found at a container.
Both can be found at the supermarket.
Both are a food.
Both are a spread.
Peanut butters is both a form of peanut butter and soup.

Is space a shape?
Both are an area.
Both are an attribute.
A star can be found at both space and shape.
The moon can be found at both space and shape.
Vacuum can be found at both space and shape.

Is a town a shopping mall?
A store can be found at both a town and a shopping mall.
A shop can be found at both a town and a shopping mall.
A barbershop can be found at both a town and a shopping mall.
A movie theater can be found at both a town and a shopping mall.
A cinema can be found at both a town and a shopping mall.
A shopping arcade can be found at both a town and a shopping mall.
Both are a municipality.
A hardware store can be found at both a town and a shopping mall.
A fast-food restaurant can be found at both a town and a shopping mall.

Is the ocean a pool?
Water can be found at both the ocean and a pool.
Both are used for swimming.
Both are used for diving.
Both have water.
Both are a body of water.
Mud puddle is both a pool and the ocean.

Is a drawer a box?
Both are a container.
Both are a form of drawers.
Clothing can be found at both a drawer and a box.
Pillowcases can be found at both a drawer and a box.

Is a clarinet a flute?
Both are an instrument.
Both can be found at orchestra.
Both are a woodwind instrument.
Both are used for making music.
Clarinets is both a form of a clarinet and a flute.
Both are used for a member of a musical organization.
Both are used for impressing a date.

Is a desk the office?
A chair can be found at both a desk and the office.
A telephone can be found at both a desk and the office.
Both are used for work.
A pencil sharpener can be found at both a desk and the office.
Staples can be found at both a desk and the office.
A lamp can be found at both a desk and the office.
A staple remover can be found at both a desk and the office.
A paperclip can be found at both a desk and the office.
Clutter can be found at both a desk and the office.
A stamp pad can be found at both a desk and the office.
Paperwork can be found at both a desk and the office.
A cup of coffee can be found at both a desk and the office.
A keyboard can be found at both a desk and the office.
A mouse can be found at both a desk and the office.
A phone can be found at both a desk and the office.
Both are used for working.
Carpet can be found at both a desk and the office.
A paper punch can be found at both a desk and the office.
A calculator can be found at both a desk and the office.
A post-it note can be found at both a desk and the office.
A desk drawer can be found at both a desk and the office.
A clock can be found at both a desk and the office.
A tape dispenser can be found at both a desk and the office.
A staple remover can be found at both a desk and the office.
Both are used for getting some work done.

Is a pocket a purse?
A pen can be found at both a pocket and a purse.
Keys can be found at both a pocket and a purse.
Money can be found at both a pocket and a purse.
A wallet can be found at both a pocket and a purse.
A key can be found at both a pocket and a purse.
A penny can be found at both a pocket and a purse.
Lint can be found at both a pocket and a purse.
A comb can be found at both a pocket and a purse.
A dollar can be found at both a pocket and a purse.
A cash coin can be found at both a pocket and a purse.
A checkbook holder can be found at both a pocket and a purse.
Both are a pouch.
A dollar bill can be found at both a pocket and a purse.

Table 2. Templates for explanation sentences

Start	
Relation	**Template**
FormOf	<FormOf> is both a form of <Word1> and <Word2>
IsA	<IsA> is both <Word1> and <Word2>
PartOf	<PartOf> is both a part of <Word1> and <Word2>
HasA	<HasA> has both <Word1> and <Word2>
UsedFor	<UsedFor> can be used for both <Word1> and <Word2>
CapableOf	<CapableOf>is capable of both <Word1> and <Word2>
AtLocation	<AtLocation> can be found at both <Word1> and <Word2>
Causes	<Causes> causes both <Word1> and <Word2>
CreatedBy	<CreatedBy> is created by both <Word1> and <Word2>
LocatedNear	<LocatedNear> is located near both <Word1> and <Word2>
MadeOf	<MadeOf> is made of both <Word1> and <Word2>
End	
Relation	**Template**
FormOf	Both are a form of <FormOf>
IsA	Both are <IsA>
PartOf	Both are a part of <PartOf>
HasA	Both have a <HasA>
UsedFor	Both are used for <UsedFor>
CapableOf	Both are capable of <CapableOf>
AtLocation	Both can be found at <AtLocation>
Causes	Both cause <Causes>
CreatedBy	Both are created by <CreatedBy>
LocatedNear	Both are located near <LocatedNear>
MadeOf	Both are made of <MadeOf>

A subway pass can be found at both a pocket and a purse.
An eyeglasses case can be found at both a pocket and a purse.
A credit card wallet can be found at both a pocket and a purse.

Is using a television surfing the net?
Both cause eye strain.
Both are used for being entertained.
Both are used for passing the time.
Both are used for entertainment.
Both are used for getting information.
Both are used for entertainment purposes.
Both cause eyestrain.
Both are used for learning.

B.1 Formatting the explanation

The explanation that is generated along with a conversation starter consists of a list of sentences, where each sentence corresponds to a matching relation found between the two words. For every possible relation a template was made, where the two words can be filled in. The templates can be found in Table 2. We make a distinction between whether the found word is at the start of a relation (***bread*** is PartOf *hot dog*) or at the end (*hot dog* can be found AtLocation ***supermarket***). The newly found word is filled into the place of the relation name, and if the template requires it, the two words in question are filled in as well. Every ConceptNet node has a 'label' which is a more completely phrased version of the word, often including an article. This helps produce a more human-readable explanation.