

Fairytale Child Chatbot

Rudolf Rosa

Charles University in Prague, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, Czech Republic,
rosa@ufal.mff.cuni.cz,
WWW home page: <http://ufal.mff.cuni.cz/rudolf-rosa>

Abstract: Fairytale Child (Pohádkové dítě) is a simple chatbot trying to simulate a curious child. It asks the user to tell a fairy tale, often interrupting to ask for details and clarifications. However, it remembers what it was told and tries to show it if possible. The chatbot can communicate in Czech and in English. It analyzes the morphology of each sentence produced by the user with natural language processing tools, tries to identify potential questions to ask, and then asks one. A morphological generator is employed to generate correctly inflected sentences in Czech, so that the resulting sentences sound as natural as possible.

1 Introduction

One of the first chatbots, and the first one that became widely known and successful, was ELIZA [10]. ELIZA tried to simulate a Rogerian psychotherapist, i.e. a therapist who tries to encourage clients to realize by themselves in what ways their feelings and behaviour are affected negatively [2]. This was a particularly good choice, as the Rogerian dialogue techniques involve asking the client general questions and giving sympathetic comments, trying to motivate him to speak more and to realize things. The chatbot analyzed the user-entered sentences in a shallow way, reacted specifically to certain keywords or patterns (“I feel tired.” – “What makes you feel tired?”), and provided a more generic answer if no keyword or pattern was spotted (such as “I understand.”, “Do you really think so?” or “Can you elaborate on that?”).

Despite its simplicity, ELIZA proved to be able to successfully communicate and react naturally enough so that numerous users reported a belief that the chatbot was actually intelligent and did understand them. One of the reasons behind the success of ELIZA is the fact that when communicating with a psychotherapist, it is the client who “creates” the world about which he talks; it is thus natural that the psychotherapist does not know anything about that world (which would appear strange in other situations, such as discussing the current political situation). Another important decision was the choice of English as the communication language, which enabled employment of rather simple pattern matching instead of full-fledged linguistic analyses, and extremely simple language generation techniques; this is thanks to English being an analytical (nearly non-inflecting) language with a fixed word order.

In this work, we present a Czech language chatbot for a similar setting – the Fairytale Child. The chatbot tries to simulate a child which wants to hear a fairy tale. Similarly to the therapist scenario, it is the user who creates the entire world in which the story takes place. The child knows nothing about this world, and thus asks a lot of questions to learn something about the world, the people and objects in it, and their properties.

The behaviour of the chatbot is similar to the “Why-phase” of child development, in which the child repeatedly asks “Why?” as a reaction to virtually any sentence. However, as the “Why-phase” is rather annoying for the parents, and we do not want our chatbot to be annoying, we try to simulate an older child, which has a wider range of more specific questions, and which often does not ask any question but simply encourages the user to go on.

Similarly to several other chatbots, the Fairytale Child remembers what it was told, and does not ask for it again; instead, it tries to show the user that it has paid attention by showing off the knowledge it has acquired if the story touches the topic again.

The Fairytale Child chatbot is built using the Treex Natural Language Processing (NLP) framework [11] both for analysis of the user input and for response generation. The tools that are part of Treex, namely the tagger [9] and morphological generator [3], are vital for the implementation of the chatbot, since they enable us to alleviate the burden of handling the morphologically rich Czech language, which is one of the main obstacles preventing usage of the same simple techniques that have been repeatedly shown to perform well enough for English.

The main aim of our work is to show that when backed by a powerful and easy-to-use NLP framework, construction of a simple Czech chatbot able to produce naturally formed responses can be not only viable, but even rather straightforward.

The chatbot is being developed in two versions – *Fairytale Child* for English, and *Pohádkové dítě* for Czech. As creating a chatbot for Czech is a more interesting and challenging endeavour, *Pohádkové dítě* is the primary focus of our work; we provide the *Fairytale Child* version mainly for our work to be intelligible also for non-Czech speakers.

2 Example

Let us start by providing a real full sample of one session of a user telling the chatbot the story of The Ugly Duck-

ling [1]. The session, listed in Section 2.1, was lead in Czech; in Section 2.2, we provide its English version.

Please observe the ways in which the chatbot tries to ask questions. Also please notice how it remembers what it was told about the objects, and when they are mentioned again, it can decide to show that it has been listening by repeating what it was told. Interestingly, this leads to a “mistake in understanding” towards the end of the story, which, by accident, seems very natural and appropriate at that point.

2.1 Czech – Pohádkové dítě

This session transcript is real and unedited (except for a few corrections of user’s typos). Therefore, it also shows several cases where the chatbot misanalyzed the sentence; these will be referred to in Section 4.

- (1) CHATBOT: Povídej mi pohádku!
- (2) USER: Povím ti pohádku o ošklivém káčátku.
- (3) CHATBOT: Bylo hodně ošklivé?
- (4) USER: No vlastně ne, ale myslelo si to o sobě.
- (5) CHATBOT: Pokračuj...
- (6) USER: Byla jednou jedna kachna, a ta nakladla vajčka.
- (7) CHATBOT: Jaká to byla vajčka?
- (8) USER: No kachní vajčka.
- (9) CHATBOT: Aha, kachní vajčka.
- (10) USER: A když přišel čas a vylhla se káčátka, jedno z nich bylo ošklivější než ta ostatní.
- (11) CHATBOT: Jaké to bylo káčátko?
- (12) USER: To káčátko, o kterém ti povídám.
- (13) CHATBOT: Aha, to káčátko, o kterém mi povídáš.
- (14) USER: Ostatní káčátka i máma kachna se mu pořád posmívali, a ono bylo z toho smutné.
- (15) CHATBOT: Byla hodně smutná?
- (16) USER: Hodně, protože ho nikdo neměl rád.
- (17) CHATBOT: Pokračuj...
- (18) USER: Pak na zimu kachny odletěly do teplých krajin, a káčátko zůstalo samo.
- (19) CHATBOT: Jaké to bylo káčátko?
- (20) USER: To ošklivé.
- (21) CHATBOT: Pokračuj...
- (22) USER: Byla mu zima, ale vzal ho k sobě domů hodný pán.
- (23) CHATBOT: Jaký to byl pán?
- (24) USER: Takový starý.
- (25) CHATBOT: Co je to starý?
- (26) USER: To je, že je mu hodně let.
- (27) CHATBOT: Pokračuj...
- (28) USER: A u toho pána káčátko přečkalo zimu.
- (29) CHATBOT: Co je to zima?
- (30) USER: To je, když je venku sníh.
- (31) CHATBOT: Jaký to byl sníh?
- (32) USER: Bílý a studený :-)
- (33) CHATBOT: Pokračuj...
- (34) USER: No a na jaře vyrazilo ven a šlo k jezeru.
- (35) CHATBOT: Jaké to bylo jezero?
- (36) USER: To, u kterého se káčátko narodilo.
- (37) CHATBOT: Pokračuj...
- (38) USER: No a káčátko uvidělo, jak k jezeru přilétají z teplých krajin krásní velcí bílí ptáci.
- (39) CHATBOT: Byli hodně bílí?
- (40) USER: Byli bílí jako sníh.
- (41) CHATBOT: Pokračuj...
- (42) USER: Byly to labutě.
- (43) CHATBOT: Pokračuj...
- (44) USER: Labutě přistávaly na jezero, a káčátko k nim opatrně šlo a obdivovalo je.
- (45) CHATBOT: Pokračuj...
- (46) USER: Čekalo, kdy si ho všimnou a budou na něj zase pokřikovat.
- (47) CHATBOT: Pokračuj...
- (48) USER: Ale labutě na něj ani nepokřikovaly, ani ho neodháněly, ani se mu nesmály.
- (49) CHATBOT: Pokračuj...
- (50) USER: Káčátko se tomu divilo, a ptalo se labutí: Copak vám nepřipadám ošklivé?
- (51) CHATBOT: Pokračuj...
- (52) USER: A na to se labutě divily a ptaly se ho: Proč bys nám připadalo ošklivé? Vždyť jsi stejně krásné jako my!
- (53) CHATBOT: Jaké to byly labutě?
- (54) USER: Ty, co na zimu odlétly do teplých krajin, stejně jako divoké kachny.
- (55) CHATBOT: Byla hodně divoká?
- (56) USER: Ale to se jen tak říká, divoké kachny, to znamená, že žijou volně v přírodě.
- (57) CHATBOT: Jaká to byla příroda?
- (58) USER: No příroda kolem toho jezera.
- (59) CHATBOT: Aha, to, u kterého se káčátko narodilo.
- (60) USER: Ano to. No a to káčátko se divilo: To přece není pravda, vy jste překrásné labutě a já jsem ošklivé káčátko.
- (61) CHATBOT: Pokračuj...
- (62) USER: A labutě mu řekly: tak pojď blíž a podívej se na svůj odraz na hladině.
- (63) CHATBOT: Jaká to byla hladina?
- (64) USER: Hladina toho jezera.
- (65) CHATBOT: Pokračuj...
- (66) USER: A tak se káčátko podívalo, ale místo ošklivého káčátka uvidělo krásnou bílou labuť.
- (67) CHATBOT: Aha, ty co na zimu odlétly do teplých krajin stejně jako divoké kachny.
- (68) USER: Právě že ne – uvidělo totiž samo sebe.
- (69) CHATBOT: Pokračuj...
- (70) USER: Nebylo to totiž ve skutečnosti káčátko, ale labuťátko.
- (71) CHATBOT: Pokračuj...
- (72) USER: No a to je konec.
- (73) CHATBOT: To už je konec? Tak děkuju za pohádku! :-)

2.2 English – Fairytale Child

This is the English version of the sample Ugly Duckling session from Section 2.1. The user utterances have been translated from Czech to English. From the responses that the English version of the chatbot is able to generate, we chose the ones that correspond to the responses in the Czech session. Thus, this session is not a real one, but it is a realistic one, in the sense that the system is able (and reasonably likely) to generate exactly the responses shown here.

- (1) CHATBOT: Tell me a fairy tale!
- (2) USER: I will tell you the tale of the Ugly Duckling.
- (3) CHATBOT: Very ugly?
- (4) USER: Well, not really ugly, but it thought so about itself.
- (5) CHATBOT: Go on...
- (6) USER: Once upon a time, there was a duck, and the duck laid some eggs.
- (7) CHATBOT: What eggs were they?
- (8) USER: Well, duck eggs.
- (9) CHATBOT: I see, duck eggs.
- (10) USER: And then when the time came and the ducklings hatched, one of them was uglier than the others.
- (11) CHATBOT: What duckling was it?
- (12) USER: Oh, the one I am telling you about.
- (13) CHATBOT: I see, the one you are telling me about.
- (14) USER: The other ducklings and the mother duck kept laughing at it, and so it was sad about it.
- (15) CHATBOT: Very sad?
- (16) USER: Very sad, because nobody liked it.
- (17) CHATBOT: Go on...
- (18) USER: Then the ducks flew south for the winter, and the duckling stayed alone.
- (19) CHATBOT: What duckling was it?
- (20) USER: Oh, it was the ugly one.
- (21) CHATBOT: Go on...
- (22) USER: It was cold, but a kind man took it into his home.
- (23) CHATBOT: What man was it?
- (24) USER: An old one.
- (25) CHATBOT: What is an old?
- (26) USER: This means that he has lived for many years.
- (27) CHATBOT: Go on...
- (28) USER: And the duckling stayed at this man throughout winter.
- (29) CHATBOT: What is a winter?
- (30) USER: This is when there is snow falling.
- (31) CHATBOT: What snow was it?
- (32) USER: White and cold :-)
- (33) CHATBOT: Go on...
- (34) USER: Well and in spring it came out and went to the lake.
- (35) CHATBOT: What lake was it?
- (36) USER: It was the lake where the duckling was born.
- (37) CHATBOT: Go on...
- (38) USER: Well, and the duckling saw beautiful huge white birds arriving to the lake from the south.
- (39) CHATBOT: Very white?
- (40) USER: They were white as the snow.
- (41) CHATBOT: Go on...
- (42) USER: They were swans.
- (43) CHATBOT: Go on...
- (44) USER: The swans were landing on the lake, and the duckling approached them carefully to admire them.
- (45) CHATBOT: Go on...
- (46) USER: It expected them to notice it and to start heckling it.
- (47) CHATBOT: Go on...
- (48) USER: But the swans did neither heckle it, nor draw it away, nor laugh at it.
- (49) CHATBOT: Go on...
- (50) USER: The duckling was surprised, and asked the swans: Don't you find me ugly?
- (51) CHATBOT: Go on...
- (52) USER: And now the swans were surprised and asked it: Why would we find you ugly? After all, you are as beautiful as we are!
- (53) CHATBOT: What swans were they?
- (54) USER: The swans that flew south, just as the wild ducks did.
- (55) CHATBOT: Very wild?
- (56) USER: Oh, this is just something you say – wild ducks means that the ducks live freely in the nature.
- (57) CHATBOT: What nature was it?
- (58) USER: Well, it was the nature around the lake.
- (59) CHATBOT: I see, it was the lake where the duckling was born.
- (60) USER: Yes, this one. Well, and the duckling was startled: This is simply nonsense, you are beautiful swans and I am just an ugly duckling.
- (61) CHATBOT: Go on...
- (62) USER: And the swans told it: well then come closer and look at your reflection at the surface.
- (63) CHATBOT: What surface was it?
- (64) USER: The surface of the lake.
- (65) CHATBOT: Go on...
- (66) USER: And so the duckling looked, but instead of an ugly duckling it saw a beautiful white swan.
- (67) CHATBOT: I see, the swans that flew south, just as the wild ducks did.
- (68) USER: Well, actually not – it actually did see itself.
- (69) CHATBOT: Go on...
- (70) USER: It wasn't in fact a duckling, but a swanling.
- (71) CHATBOT: Go on...
- (72) USER: And this is the end.
- (73) CHATBOT: This is the end already? Thanks for the story then! :-)

Czech sentence analysis scenario
Treex::Block::W2A::CS::Tokenize
Treex::Block::W2A::CS::TagMorce
Treex::Block::W2A::CS::FixMorphoErrors
English sentence analysis scenario
Treex::Block::W2A::EN::Tokenize
Treex::Block::W2A::EN::NormalizeForms
Treex::Block::W2A::EN::FixTokenization
Treex::Block::W2A::EN::TagMorce
Treex::Block::W2A::EN::FixTags
Treex::Block::W2A::EN::Lemmatize

Table 1: Treex scenarios for sentence analysis.

3 Implementation

The chatbot is implemented as a console application in Perl, reading user input from standard input, and producing responses to standard output. It extensively uses Treex framework blocks and tools, but is not implemented as a Treex application, as Treex scenarios do not support interactive operation.

The script starts by initializing all tools and asking the user to tell a fairy tale (see utterance (1) in Section 2). Then it iteratively analyses the sentence entered by the user with NLP tools (Section 3.1), selects a suitable predefined response template and generates a response (Section 3.2).¹ This loop continues until the end of the story is detected – this is done by detecting the presence of the word “end” (“konec” in Czech), see (72)-(73). The chatbot thanks for the story and terminates.

3.1 User Utterance Analysis

Each line entered by the user is analyzed with a standard Treex scenario – see Table 1. The sentences are split into individual tokens, and each token is assigned a lemma² and a part-of-speech³ tag.

For Czech, the standard Czech positional tagset is used [5]. The tag marks several morphological categories; in our work, we make use of the following ones: part-of-speech, gender, number, case, person. An important feature of the tagset is that while one word form may often be analyzed in several ways due to inflection ambiguity (e.g. “staré”, “old”, may be e.g. neuter nominative singular, feminine genitive singular or feminine accusative plu-

¹Our chatbot, as most other chatbots, generates a response to each user input. However, it would be very annoying if the chatbot generated a question each time – unlike Eliza, our system expects a user who knows what he wants to say and does not need to be asked something all the time to elicit input from him; we want to simulate a child that actually does want to hear the story and therefore sometimes just listens, without asking anything. Instead of being completely silent, which could be mistaken for long response time or for an error, we therefore sometimes either simply encourage the user to go on, or show that the chatbot is listening by showing what it remembers from what the user had said.

²E.g. for the token “sons”, the lemma is “son”

³E.g. noun, pronoun, verb

ral), the vice versa is not possible: for a given lemma and tag, there is always at most one corresponding word form.

For English, the standard Penn Treebank tagset [6] is used. Unfortunately, it does not mark the natural gender for English nouns, so e.g. both “boy” and “girl” are tagged identically as NN. However, it does mark noun number, which is a feature that we make use of.

If the chatbot asks a question, the user’s answer is stored in a simple knowledge base. Before storing, the answer is normalized in two steps. The first one is the removal of sentence-initial garbage words “Well”/“Oh” (“No” in Czech) – see (8)-(9). The second one is more intricate: it is a switch between 1st and 2nd person – see (12)-(13). This consists of switching “I”/“me” and “you” in English (“já” and “ty” in Czech), and also switching the person for Czech pronouns and verbs.

The knowledge base is simply a key-value hashmap. The key is the lemma of the word (noun or adjective) that the chatbot asked about, and the value is the normalized answer of the user to that question.

3.2 Response Selection and Generation

The chatbot randomly chooses from a set of the following three possible response types. It then tries to generate the response of the given type. If this is impossible, e.g. because the template requires a noun but there is no noun in the sentence – see (5) – it responds “Go on...” (“Pokračuj...”) instead. This randomness makes the responses less predictable and therefore more realistic.

For generating Czech responses, a morphological generator [3] is used to generate appropriately inflected word forms based on (lemma, tag) pairs.

What NOUN was it? The first response type asks the user to provide details about a noun – see e.g. (23). The noun to be asked about (denoted as NOUN) is the last noun in the sentence, as this is presumably the newest and most interesting piece of information.

In Czech, NOUN is first switched to nominative case using the morphological generator (but morphological number is preserved). The response generated has the form of “*Jak_ to byl_ NOUN?*”; the correct form of “*Jak_*” (“*Jaký*”, “*Jaká*”, “*Jaké*”... – “What”) and “*byl_*” (“*byl*”, “*byli*”... – “was”/“were”) is chosen based on the morphological gender and number of the NOUN (and generated by the morphological generator).

In English, the response has a form of “What NOUN was it/were they?”, based on the singular/plural marked in the NOUN tag (NN and NNP tags mark singular, NNS and NNPS tags mark plural).

If a question regarding NOUN had already been asked (i.e. the knowledge base already contains NOUN as a key), the chatbot does not ask the question again, but instead shows off by repeating the then-provided answer – “I see,

ANSWER.” (“Aha, ANSWER.”) – see (35)-(36) and then (58)-(59).⁴

What is a NOUN? The second response type asks the user to explain what a noun means – see e.g. (29). Again, the noun to be asked for (denoted as NOUN) is the last noun in the sentence.

In Czech, NOUN is first switched to nominative, preserving the number. The response generated has the form of “Co je/jsou to NOUN?”, the correct choice of verb “je” or “jsou” (“is” or “are”) being made based on the morphological number of the noun.

In English, the response has a form of “What is/are a/an NOUN?”, based on the singular/plural marked in the NOUN tag, and on the first letter of NOUN.⁵

Similarly to the first response type, if a question regarding NOUN had already been asked, the chatbot repeats the answer instead.

Very ADJECTIVE? The third type asks whether a property was intense. Similarly to previous types, the adjective to be asked about (denoted ADJECTIVE) is the last adjective of the sentence.

In Czech, the ADJECTIVE is first converted to nominative. Then, the response is generated in the form “Byl_hodně ADJECTIVE?” (the correct form of “byl_” is generated according to morphological gender and number of ADJECTIVE).

In English, the form is only “Very ADJECTIVE?”, for reasons discussed in Section 4.

If a question regarding ADJECTIVE had already been asked, the chatbot resorts to the default “Go on” answer, as for adjective questions, the answers are not reproduced. The idea is that while it is reasonable to believe that a repeated noun is a reference to the same entity, this is less likely for adjectives. However, to avoid annoying questions, the chatbot will not ask about the same adjective twice.

3.3 Efficiency Evaluation

Although both Treex and Perl are sometimes blamed for their slowness, we have not encountered such issues with our application. This is helped by the fact that all the Treex tools are loaded only once – thus, the startup takes about 7

⁴Note that (10)-(11) and (18)-(19) seem to contradict this claim. However, on closer inspection of the original Czech transcript, it can be seen that the user used different spellings for the word “duckling” – “káčátko” in (10) but “kačátko” in (18). This difference has no influence on meaning in Czech, but as the tagger preserves this distinction in the lemmas it produces, our system treats the two as distinct, and therefore asks about the duckling again instead of showing that it already knows what the user is talking about. (The distinction has not been preserved in the English version because of the lack of such a dublet in English.)

⁵We resort to a simplification here by treating all vowel letters as vowels – i.e. we would incorrectly generate e.g. “an uniform” instead of the correct “a uniform”.

seconds for Czech (about 3 seconds for English), but each response then takes only about 70 ms for Czech (about 15 ms for English).⁶ We believe this to be perfectly acceptable for real-world usage.

4 Discussion and Future Work

The whole conversation between the user and the chatbot has a reasonably natural feel. The chatbot usually reacts appropriately, which is demonstrated by the ability of the user to answer chatbot’s questions and continue with the story, eventually reaching the end.

Occasionally, we encounter errors that stem from incorrect morphological analysis disambiguation by the tagger. One such case can be seen in (15), where “Byla hodně smutná?” (“Was she very sad?”) was generated instead of the correct “Bylo hodně smutné?” (“Was it very sad?”), as the neuter nominative “smutné” (“sad”) in (14) was mis-analyzed as feminine genitive. A similar situation happens also in (54)-(55), where feminine plural nominative is mis-analyzed as feminine singular genitive, and in (24)-(25), where “starý” (“old”) is analyzed as a noun instead of adjective. However, such errors happen already in the tool that we use, and are effectively beyond our influence.

An error that we have encountered (but does not show in the sample conversation) is that of asking the “Very ADJECTIVE?” question with inappropriate adjectives, as in “He sat there the whole day.” (“Seděl tam celý den.”) – “Very whole?” (“Velmi celý?”). However, it seems to occur especially with several frequent adjectives, so enumerating them manually seems to be a viable remedy.

We also experimented with the “Why?” question, but have found it hard to identify the appropriate part of the user’s utterance to use in the question – the whole sentence seems often too long and awkward, while only taking the verb often leads to a non-sentence. This is an issue to be addressed once we add deeper levels of linguistic analyses.

There are still issues with the English implementation. One of them is that the Czech question with “Jaký”, such as (23) “Jaký to byl pán?”, is conveniently ambiguous, as it can mean both “What was the man like?” and “Which man was it?”, and thus feels more natural than the English one. In English, we are not aware of such an ambiguous construction, and we thus use “What man was it?”, which tries to keep the ambiguity but is not good English.

There are also cases where the Czech rich morphology and analysis helps – especially when dealing with gender, which is not marked in English analysis. Thus, when asking the “Very ADJECTIVE?” question, we can choose the correct subject in Czech based on the morphological features of the adjective, while this is not possible in English. This can be seen e.g. in (3), where in Czech we generate the whole question “Bylo hodně ošklivé?” (“Was it

⁶Measured on Intel® Core 2 Duo T5870 running 32b Ubuntu 10.10, averaged over 5 runs. The response time is averaged over the 36 sentences from Section 2.

very ugly?”), because the morphological tag tells us that “ošklivé” (“ugly”) was used in a neuter gender, while in English, we currently resort only to “Very ugly?”, since we cannot easily distinguish whether we should use “he”, “she” or “it” (or another pronoun) as the subject. This is often distractive and leads to unnatural questions, as in (39). The gender is not only unmarked on the adjective itself, but the English tagset does not even mark gender on nouns (so there is no simple and direct way of telling that a “king” is a “he” while a “queen” is a “she”). However, there exist tools that are able to fill that shortcoming, and we will try to use them in future. Still, it will necessitate to use not only tagging but also syntactic parsing to be able to identify the gender associated with an adjective, since it is implied by the parent noun in a dependency tree, which might not be straightforward to find without a proper syntactic analysis of the sentence.

Employing a syntactic parser and a function labeller, such as [7], would be beneficial in other ways as well, e.g. by enabling us to use meaningful phrases in the responses, as opposed to the current approach where we can only use either a single word or the whole sentence, or by being able to distinguish the syntactic subject from the syntactic object. An analysis of the topic-focus articulation [4] could probably make it possible to generate even more natural questions, trying to concentrate on questions about the *focus*, i.e. the new information – not about the *topic*, which is the information that the speaker assumes is already known to the listener.

An obvious and important future goal is to implement a web interface. Currently, the application is terminal-based, reading user utterances from the standard input and producing chatbot responses to standard output. This does not look very appealing; moreover, it requires the users to have Treex installed. Even a simple web interface would make the application available to a much wider range of users, and would also make it more attractive. While Treex does have a web interface [8], the Treex scenarios are not designed for interactive use, and therefore a different solution will have to be found.

5 Conclusion

We have implemented a Fairytale Child chatbot, which acts as a child that wants to hear a fairy tale. It analyzes the user utterances and tries to react with a question or remark if possible, simulating a curious child. The chatbot is able to communicate in Czech or in English.

We used the Treex natural language processing framework for both analysis of the user utterances and generation of the chatbot responses. This proved to be effective, especially for the morphologically rich Czech language.

Even though the implementation of the chatbot is quite simple, it responds quite naturally in most cases. This shows that even for morphologically rich languages, a simple approach can lead to a well-performing chatbot, provided that there is a powerful NLP framework at hand.

The chatbot is licenced under GNU GPL v2 and is available at <https://ufal.mff.cuni.cz/dite>.⁷

Acknowledgments

This research was supported by the grants GAUK 1572314, and SVV 260 104. This work has been using language resources developed, stored and distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2010013).

References

- [1] Hans Christian Andersen. *New Fairy Tales*. C. A. Reitzel, Copenhagen, 1844.
- [2] Lisa M. Cepeda and Donna S. Davenport. Person-centered therapy and solution-focused brief therapy: An integration of present and future awareness. *Psychotherapy: Theory, Research, Practice, Training*, 43(1):1, 2006.
- [3] Jan Hajič. *Disambiguation of Rich Inflection (Computational Morphology of Czech)*. Nakladatelství Karolinum, 2004.
- [4] Eva Hajičová. Topic-focus articulation and coreference in models of discourse production. *Journal of Pragmatics*, 16(2):157–166, 1991.
- [5] Jan Hajič. Building a Syntactically Annotated Corpus: The Prague Dependency Treebank. In Eva Hajičová, editor, *Issues of Valency and Meaning. Studies in Honor of Jarmila Panevová*, pages 12–19. Prague Karolinum, Charles University Press, 1998.
- [6] Mitchell P. Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. Building a large annotated corpus of English: the Penn treebank. *Comp. Ling.*, 19:313–330, June 1993.
- [7] Ryan McDonald, Fernando Pereira, Kiril Ribarov, and Jan Hajič. Non-projective dependency parsing using spanning tree algorithms. In *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 523–530, Vancouver, British Columbia, Canada, 2005.
- [8] Michal Sedlák. *Treex::Web*. Bachelor’s thesis, Charles University in Prague, Faculty of Mathematics and Physics, Prague, Czechia, 2014.
- [9] Drahomíra Spoustová, Jan Hajič, Jan Votrubec, Pavel Krbec, and Pavel Květoň. The best of two worlds: Cooperation of statistical and rule-based taggers for Czech. In *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing, ACL 2007*, pages 67–74, Praha, 2007.
- [10] Joseph Weizenbaum. Eliza – a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.
- [11] Zdeněk Žabokrtský. Treex – an open-source framework for natural language processing. In Markéta Lopatková, editor, *ITAT*, volume 788, pages 7–14, Košice, Slovakia, 2011. Univerzita Pavla Jozefa Šafárika v Košiciach.

⁷The 1.0 version of the chatbot, which is described in this paper, is published at: <http://hdl.handle.net/11234/1-161>