Toward Robust Semantic Interaction for English Language Learners

Master's Thesis

Presented to
The Faculty of the Graduate School of Arts and Sciences
Brandeis University
Department of Computer Science
Nianwen Xue, Advisor

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
in
Computational Linguistics

by
Trung Luu

May 2016
ACKNOWLEDGEMENTS

I want to dedicate this thesis to all of my teachers, friends and students who are lifelong language learners and educators, especially my parents, who first showed me the beauty of foreign languages and cultures.

I am very grateful to James Pustejovsky, who shaped my view on semantics as a dynamic ecosystem evolving generatively and creatively.

My deepest gratitude goes to my advisor Nianwen (Bert) Xue, who first introduced me the topics of Abstract Meaning Representation and Coreference Resolution. Thank you, Bert, for your generous approval of my vague ideas, and then making them clear and grounded by your thought-provoking questions and suggestions. Your instant and straightforward feedback always gives me timely clarifications and directions in my work and develops me into a better critical thinker.

I owe a debt of gratitude to Sophia A. Malamud and Marie Meteer, who have provided me with plenty of unique and precious opportunities to immerse into compelling discussions on discourse and pragmatics, including discourse coherence, entity salience, Centering theory, conversational scoreboards and dialog systems, from both theoretical and computational perspectives.

I would like to thank to my fellows in our Computational Linguistics program for all of their input and encouragement that have definitely boosted my confidence and performance throughout this work.

I am very grateful to all of my committee members, especially Lotus Goldberg, my favorite mentor, whose appearance in our defenses is an extremely sweet surprise to me,
and Keith Plaster, whose diligent support makes my defense one of the best experiences I have had as a master student at Brandeis.

I also want to thank all of those who have helped me in the smooth transition to and integration in Brandeis community, including my family, our Master’s program in Computational Linguistics, Computer Science Department, Writing Center, Desktop Computing Division of Library & Technology Services, Graduate Student Affairs, Graduate School of Arts and Sciences, and International Students and Scholar Office. Without you, I could not survive till the end of our program and have any chance to work out this thesis.

Last but not least, I am very grateful to my lovely friends namely The Little Prince, Cinderella, Ariel, Aurora, Belle, and Snow White. Your pristine beauties always keep me fresh and tough, creating an inside voice urging me to work hard on this thesis. I do wonder if I was able to finish it in time without you.
ABSTRACT

Toward Robust Semantic Interaction for English Language Learners

A thesis presented to the Department of Computer Science
Graduate School of Arts and Sciences
Brandeis University
Waltham, Massachusetts

By Trung Luu

Facing the real-life needs for personalized interaction capable of assisting learners in actively engaging with a variety of language resources to promote their own development, language educators have been expecting Intelligent Language Tutoring Systems (ILTS) as the ultimate solutions that capture the interdisciplinary quintessence of Human Language Technology. However, the majority of current ILTS do not address the innovative changes in language teaching methodology that supports the balance between the structural knowledge of language (form-focused) and the ability to use it for practical communication purposes (meaning-focused). Specifically, they tend to heavily focus on form-based interaction such as corrective feedback on grammatical errors. In this context, we aim to develop a text-based dialog system that is capable of acquiring semantic and pragmatic knowledge to create meaning-focused interaction for English language learners (ELLs).
Specifically, we first develop a robust semantic representation of a online authentic text that is integrated with plenteous meaningful features for language learning purposes and then design learning units as coherent sequences of individualized meaning-focused interactions between a conversational agent and a learner, based on the developed semantic representation of the core text. To address the first task, we chose the sentence-level Abstract Meaning Representation (AMR) formalism as the starting point and make an pioneer attempt at solving the problem of discourse level coreference resolution for AMR texts as sequences of AMR graphs. Inspired by Centering theory and Stanford’s rule-based coreference resolution system, the proposed algorithm features a simple cache model for searching antecedent candidates, a versatile and scalable framework for coreference feature integration, and a multiple-factor model of salience for ranking antecedent candidates. The output cross-sentential co-referred concepts are then merged together to develop the document-level AMR graph for the examined text. Based on this robust text-based semantic representation, we implement a proof-of-concept dialog system handling Semaland game, an exclusive in-house design that allows ELLs to truly interact with the system’s conversational agent regarding the semantic concepts in the text they have read, and therefore provides the learners with beneficial individualized learning experiences.
# TABLE OF CONTENTS

1. Introduction .................................................................................................................. 1

2. Rationale ......................................................................................................................... 3
    2.1. Adequate Pedagogical Paradigms ............................................................................. 3
        2.1.1. From the perspective of learning theories ....................................................... 4
        2.1.2. From the perspective of SLA .......................................................................... 4
        2.1.3. From the perspective of ELT methodology ..................................................... 5
    2.2. The status quo of ILTS ............................................................................................ 6
    2.3. Research Objectives ............................................................................................... 7

3. Text-based Semantic Representation for ELL interaction ............................................. 8
    3.1. Cinderella Corpus .................................................................................................... 8
    3.2. Sentence-Level Semantic Representation - AMR graph ......................................... 9
        3.2.1. AMR Formalism as an Optimal Option .......................................................... 10
        3.2.2. AMR Parsing of Cinderella Corpus .................................................................. 11
    3.3. Document-Level AMR-based Coreference Resolution .......................................... 12
        3.3.1 Theoretical Framework ...................................................................................... 13
        3.3.2. Task Description ............................................................................................. 15
        3.3.3. Algorithm Description ..................................................................................... 19
        3.3.4. Experiments ...................................................................................................... 26
    3.4. Document-Level AMR-based Semantic Representation .......................................... 28
        3.4.1. Definition of Linking Concepts between Sentence-Level AMR Graphs ............. 29
        3.4.2. Merging of Sentence-Level AMR Graphs into Document-Level AMR Graph ....... 30

4. Semaland Game - Proof-of-Concept Dialog System for ELL interaction ..................... 33
    4.1. Description of Semaland Game ............................................................................... 33
    4.2. Implementation of Semaland Game Dialog System ............................................... 36
    4.3. Demonstration of Semaland Game Interaction ....................................................... 38

5. Conclusion and Future Work ....................................................................................... 40

References .......................................................................................................................... 42
LIST OF TABLES

Table 1: Preliminary statistics of Cinderella corpus.............................................................. 9
Table 2: Evaluation of the existing AMR parsers..................................................................... 11
Table 3: Feature effects (cache size = 2)................................................................................. 26
Table 4: Cache size effects........................................................................................................ 27
Table 5: Salience factor effects (cache size = 1)..................................................................... 27
Table 6: Evaluation result on the test set (cache size = 1)....................................................... 28
Table 7: AMR-based statistics of Cinderella corpus................................................................. 29
Table 8: Merging algorithm for document-level AMR graph.................................................. 31
LIST OF FIGURES

Figure 1: Example of an AMR graph................................................................. 12
Figure 2: AMR graph of a sentence in *The Little Prince* corpus.......................... 18
Figure 3: Cache model..................................................................................... 20
Figure 4: AMR graph including “thing” nodes.................................................... 23
Figure 5: Various strategies for coreference resolution....................................... 24
Figure 6: Document-level AMR graph of *Cinderella* story................................. 32
Figure 7: *Semaland* Dialog Manager................................................................ 37
Figure 8: *Semaland* game demonstration........................................................ 39
1. Introduction

Research in Natural Language Processing (NLP) field can be grouped in various dimensions such as the aspects of language knowledge (e.g. phonetics, phonology, morphology, syntax, semantics, pragmatics and discourse), the nature of applications (e.g. machine translation, information extraction, question answering, summarization, dialog systems and conversational agents), and the domains of application (e.g. market research, customer service, device management, education and entertainment) (Jurafsky & Martin, 2009). From this multidimensional perspective, this work aims to develop a text-based dialog system that is capable of acquiring semantic and pragmatic knowledge to create meaning-focused interaction for English language learners (ELLs). It is worth noting that unlike the mainstream research in Artificial Intelligence (AI) field that focuses on the machine capability, this work primarily serves the human communicative competence, and thus seeks for a robust solution to the optimal integration of Human Language Technology (HLT) into the available learning resources for world-wide ELLs.

Needless to say, among recent technological developments, the Internet becomes the single invention with the greatest impact on language teaching and learning (Warschauer, 1996) as it facilitates “integrative interaction with authentic learning materials across a wide spectrum of virtual contexts” (Krummes, 2013) in an extremely flexible manner which opens a huge space for learner creativity, both in terms of learners’ selecting tools according to their own needs and in sense of linguistic and stylistic
creativity. However, the easy access to the tremendous resources on the Internet poses another challenge for ELLs because having a great deal of learning options, they tend not to engage with a specific material deeply enough to achieve a complete learning experience with a high knowledge retention rate. From this perspective, this work proposes a robust solution to deepening ELLs’ engagement with online language contents for their better learning experiences.

To best describe the contents of the current work, this thesis is organized as follows. Chapter 2 presents a detailed discussion on the rationale of this work which can be skipped by the readers who are more interested in its technical contents. This discussion helps us gain strong confidence that this work addresses relevant research problems and possesses solid theoretical foundations. Chapter 3 addresses the key technical concern of this work namely the robust text-based semantic representation for ELL interaction. Specifically, this chapter includes a relatively independent and comprehensive discussion on the document-level coreference resolution for Abstract Meaning Representation (AMR) graphs (Banarescu et al., 2013), which, to the best of my knowledge, has not been directly examined in any published work. Chapter 4 provides a simplistic framework for a dialog system that capitalizes on the text-based semantic representation developed in Chapter 3 in order to create meaning-focused interaction for ELLs. The designed communication is named Semaland game, reflecting its Formal Pragmatics and Game Theory based origin. Chapter 5 concludes and makes some key suggestions for future work. Finally, it is worth mentioning that the solutions presented in this work vividly demonstrate my intentional efforts to combine the strengths of both rationalist and empiricist approaches in the realm of scientific research (Olson & Hergenhahn as cited in Olson, 2015, p. 13).
2. Rationale

In modern Second Language Education (SLE) there are real-life needs for personalized interactions including individual feedback and guidance which effectively raise learners’ awareness of relevant language phenomena as well as assist them in actively engaging with a variety of language resources to promote their own development (Meurers, 2009). To satisfy these critical needs of a steadily growing population of worldwide language learners, taking into account the various constraints on in-class resources, language educators have been expecting Intelligent Language Tutoring Systems (ILTS) as the ultimate solutions which capture the interdisciplinary quintessence of NLP and SLE and therefore are supposed to be capable of delivering comprehensive learning experience to language learners. From this perspective, this chapter discusses the instructional desiderata for ILTS, the shortcomings of state-of-the-art ILTS and the research objectives of the current work.

2.1. Adequate Pedagogical Paradigms

In order to develop robust ILTS for real-life needs, it is sensible to start with the building of an compelling pedagogical paradigm derived from well-established research findings in learning theories, Second Language Acquisition (SLA) and English Language Teaching (ELT) methodology.
2.1.1. From the perspective of learning theories

It is clear that every serious learning and teaching system should be developed from certain learning theories whose classic embodiment includes behavioral, cognitive, and socio-cognitive frameworks (Sistelos, 2008). Among them sociocultural theory, originally proposed by Vygotsky, is one of the most prominent and influential approaches to the state-of-the-art SLE. The essence of this theory is the concept of Zone of Proximal Development (ZPD) which sheds the light on the socio-cultural context in which learners’ cognitive development occurs and grows out of their interactions with members of their culture. Recently, this concept is compellingly expanded to reflect the process of self-scaffolding or ZPD management: learners promote their own development by proactively engaging in meaningful interactions with appropriate learning resources (Bickhard as cited in Ohta, 2013). Moreover, successful learners are the ones who have higher level of engagement in subsequent activities related to new-learned language phenomena, or the ones who have deeper learning experience (Knouzi et al. as cited in Ohta, 2013). From practical perspective, the clearest example of this process is nowadays’ web-based exploratory learning.

2.1.2. From the perspective of SLA

One of the most widely known and influential SLA hypotheses among language educators is Krashen’s Acquisition-Learning distinction. According to Krashen, acquisition is a subconsciously process which is very similar to the process that children undergo when learning their native language; meanwhile, learning is a conscious process which involves
perceiving the formal instruction in traditional classrooms (Towell, 2013). Krashen asserted that acquisition which requires natural and meaningful interaction in the target language is much more effective than formal learning. Interestingly, Krashen’s emphasis on acquisition shares a compelling playground with Interaction Model in SLA in which the learning of the Second Language (L2) is claimed to evolve out of communicative use per se when learners are exposed to input, produce output and receive feedback on that output (Mayo & Soler, 2013). In this model, learners are motivated to connect a wide range of linguistic forms onto their meaning through appropriate interactions.

2.1.3. From the perspective of ELT methodology

Based on the findings of SLA studies, theory-driven ELT methodology is recently characterized by the so-call post-method/ appropriate pedagogy (Waters, 2012). In other words, the invariant of the ELT methodology is no longer a certain teaching/ learning method for how language can best be taught and learnt, but the methodical eclecticism supporting the desirable twin goals of equipping learners with a knowledge of the “system” of English (form-focused) and the ability to use it for practical communication purposes (meaning-focused). Consequently, the novel challenge is to find the best balance of diverse methodical components in language instruction which depends not only on current insights into SLA processes but also on situational factors namely learners’ background and expectation, the influence of stakeholders such as parents and policy makers, as well as different constraints on teaching and learning resources. Indeed, this challenge has been creating a stable gap between the theory and practice in modern ELT methodology in which the former focuses more on ‘communicating to learn’-oriented approach while the
latter tends to adopt more conservative ‘learning to communicate’-based style. Notwithstanding the considerable stability of this gap over decades, the revolutionarily growing use of electronic technology in language teaching is expected to signal a radical change in the coming years; for example, there is convincing evidence that the increasing ubiquity of web-based language teaching and learning resources has the potential to redistribute the balance between teacher-led and learner-based interactions.

2.2. The status quo of ILTS

Reviewing current ILTS (Meurers, 2012), we can find the following limited adequateness of the majority of these systems in current SLE: they have been developed either for the analysis of learner language in order to give the corrective feedback on form and, more rarely, content errors made by learners as well as score learners’ work automatically, or for the analysis of native language in order to retrieve authentic language learning resources. In fact, none of these focuses directly address the innovative changes in SLA and ELT methodology surveyed above.

These shortcomings are actually addressed in Lee et al., (2015), one of the most recent attempts to balance the situation. According to these authors, “previous dialog-based language learning systems mostly only play the role of a conversational partner using chatting like spoken dialog technology, and providing feedback such as grammatical error correction and suggesting better expressions.” In contrast, they develop a conversational educational teaching agent integrated with a knowledge base for the capability of maintaining students’ interest during their learning process through meaning-focused interaction. Needless to say, this pioneering work clearly signals a new wave of
NLP research for SLE purposes that is in favor of the ‘communicating to learn’-oriented approach.

2.3. Research Objectives

This work is the first step of an original approach to develop ILTS which can effectively integrate into the state-of-the-art SLE contexts, featuring the application of NLP to the robust creation of natural and meaningful interactions for language learners within the recommended ELT methodology which thoughtfully contribute to the optimal balance of diverse methodical components in language instruction.

Specifically, the work addresses the lack of meaning-focused interaction available in current ILTS by first developing a robust semantic representation of an online authentic text that is integrated with plenteous meaningful features for language learning purposes and designing learning units as coherent sequences of meaning-focused interactions between a conversational agent and an ELL which evolve from the developed semantic representation of the core text. As the ultimate goal is to offer learners the genuine opportunities to promote their own development - self-scaffolding - the developed semantic representation should have a rhizomatic (graph-based) structure which allows each learner has their own path of navigation through their learning experience (Lian, 2004).

Last but not least, in addition to the mainstream purpose of developing ILTS for language learners, this work can create a new environment for developing an invaluable learner corpus which is definitely beneficial for diverse interdisciplinary research.
3. Text-based Semantic Representation for ELL interaction

This chapter is the central part of this work, covering a complete development cycle of a robust text-based semantic representation that is optimal for the later ELL interaction design task. The discussions here include the decision-making in the selection of the relevant learning texts and the adequate sentence-level semantic representation as well as the solution to the development of the corresponding document-level semantic representation for the selected texts.

3.1. Cinderella Corpus

As emphasized at the beginning of this work, Internet gives learners invaluable access to tremendous authentic language materials via various (asynchronous and synchronous) communication channels such as websites, blogs, wikis, email, chat, forum, video-conferencing, social networking and virtual environments (Krummes, 2013). In addition, the language varieties in relation to internet-based interactions vividly extend the language learning and communication experience to cultural awareness and intercultural competence. Therefore, it is sensible to select online authentic and culturally enriched texts that are enhanced by multimedia input channels, such as visual and aural, for English language learning purposes as the studied objects for the current research. Understandably, fairy tales adopted by Walt Disney cartoon empire are definitely a perfect choice.
Specifically, the short versions of five stories in the collection of Walt Disney princesses, including Ariel\(^1\), Aurora\(^2\), Belle\(^3\), Cinderella\(^4\) and Snow White\(^5\), are selected for form a small corpus name *Cinderella* (as this character is usually considered the unofficial leader of Walt Disney princesses). All of these texts are greatly presented with illustrative pictures, narrative voice and background music on Walt Disney’s website and. In *Cinderella* corpus, the raw texts are saved in TXT format. Several statistics of this corpus are presented in Table 1.

<table>
<thead>
<tr>
<th>No.</th>
<th>Text name</th>
<th>Sentence count</th>
<th>Word count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ariel</td>
<td>45</td>
<td>518</td>
</tr>
<tr>
<td>2</td>
<td>Aurora</td>
<td>51</td>
<td>756</td>
</tr>
<tr>
<td>3</td>
<td>Belle</td>
<td>48</td>
<td>544</td>
</tr>
<tr>
<td>4</td>
<td>Cinderella</td>
<td>52</td>
<td>548</td>
</tr>
<tr>
<td>5</td>
<td>Snow White</td>
<td>38</td>
<td>448</td>
</tr>
<tr>
<td></td>
<td><strong>Total:</strong></td>
<td><strong>234</strong></td>
<td><strong>2814</strong></td>
</tr>
</tbody>
</table>

**Table 1**: Preliminary statistics of *Cinderella* corpus.

It is worth noting that *Cinderella* corpus (and other code-related components of this project) can be directly accessed via https://github.com/luutuntin/Cinderella.

### 3.2. Sentence-Level Semantic Representation - AMR graph

This section reviews available varieties of semantic representation and explains the adequacy of AMR formalism (Banarescu et al., 2013) for the task addressed in this work.

---

\(^1\) http://princess.disney.com/ariels-story  
\(^2\) http://princess.disney.com/auroras-story  
\(^3\) http://princess.disney.com/belles-story  
\(^4\) http://princess.disney.com/cinderellas-story  
\(^5\) http://princess.disney.com/snow-whites-story
Further, different AMR parsers are taken into consideration before the most promising one is chosen for automatically parsing *Cinderella* corpus.

### 3.2.1. AMR Formalism as an Optimal Option

It is widely recognizable that there is no universal semantic representation for every NLP application. Instead, each domain-specific application requires different computer executable complete meaning representation which is sufficient to achieve the task under consideration (Kate & Wong, 2010; Artzi et al., 2013). In the scope of this work, in order to design personalized meaning-focused interaction for ELLs, we need to represent the semantic contents of examined texts coherently (i.e. all semantic objects should link together) and comprehensively (i.e. all meaningful concepts should be captured) so that the system has the sufficient knowledge to select a relevant topic for the ongoing interaction based on the user’s most recent response and the history of the corresponding conversation. In other words, the desirable semantic representation for the addressed task should be capable of covering the complete meaning of various texts in a structural manner. Reviewing currently available formalisms of semantic representation in NLP (Liang, 2015), we can argue that AMR, the most comprehensive embodiment of frame semantics, is the best candidate which unifies all essential semantic annotation for well-known tasks such as semantic role labeling, named-entity recognition and coreference resolution.

In essence, AMR is a semantic representation language aimed at creating a giant human-annotated semantics bank (Banerescu et al., 2013). Based on neo-Davidsonian semantics, AMR captures the whole-sentence meaning of “who is doing what to whom” in the form of a rooted, directed, acyclic graph with labels on nodes (concepts) and edges
(relations between concepts). This single all-in-one data structure is very robust in the sense that it abstracts from syntax and therefore can represent various sentences that have the same basic meaning.

### 3.2.2. AMR Parsing of Cinderella Corpus

Thanks to the recent release of sizable corpora of English/AMR pairs, including LDC2013E117 and LDC2014T12, several AMR parsers are developed to address the problem of automatic AMR parsing. The Smatch\(^6\) (Cai & Knight, 2013) evaluation of the existing AMR parsers on the newswire section of these datasets is presented in Table 2.

<table>
<thead>
<tr>
<th>System</th>
<th>LDC2013E117-proxy</th>
<th>LDC2014T12-proxy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>JAMR(^7) (Flanigan et al., 2014)</td>
<td>67</td>
<td>58</td>
</tr>
<tr>
<td>MSR SPLAT(^8) (Vanderwende et al., 2015)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Stanford (Werling et al., 2015)</td>
<td>66</td>
<td>59</td>
</tr>
<tr>
<td>SHRG-based (Peng et al., 2015)</td>
<td>59</td>
<td>57</td>
</tr>
<tr>
<td>CAMR(^9) (Wang et al., 2015)</td>
<td>71</td>
<td>69</td>
</tr>
<tr>
<td>IST(^10) (Pust et al., 2015)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Cornell(^11) (Artzi et al., 2015)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Table 2**: Evaluation of the existing AMR parsers.

In spite of the modesty of these results, the application of these AMR parsers, particularly JAMR parser, immediately shows promising results in addressing such tasks as abstract summarization (Liu et al., 2015) and entity linking (Pan et al., 2015) which require

---

\(^6\) [http://amr.isi.edu/evaluation.html](http://amr.isi.edu/evaluation.html)

\(^7\) [https://github.com/jflanigan/jamr](https://github.com/jflanigan/jamr)


\(^9\) [https://github.com/Juicechuan/AMRParsing](https://github.com/Juicechuan/AMRParsing)

\(^10\) [http://www.isi.edu/~pust/](http://www.isi.edu/~pust/)

\(^11\) [https://bitbucket.org/yoavartzi/amr](https://bitbucket.org/yoavartzi/amr)
more or less deep semantic knowledge at document level. In the current work, CAMR, the best AMR parser\textsuperscript{12} according to the evaluation scores in Table 2, is used to parse Cinderella corpus with the confidence that its effectiveness can be transferable to the addressed task. Figure 1 displays an example of parsed sentences in Cinderella corpus.

![AMR Graph Example](image)

**Figure 1**: Example of an AMR graph\textsuperscript{13}.

### 3.3. Document-Level AMR-based Coreference Resolution

To develop a robust semantic representation of a whole text, we need to merge its sentence-level AMR graphs into a bigger document-level AMR graph, using cross-sentential co-referred concept nodes. In other words, we need to address the problem of discourse

\textsuperscript{12} using model \textit{LDC2013E117.train.basic-abt-charniak.m}

\textsuperscript{13} visualized by amr-reader toolkit (Pan et al., 2015)
level coreference resolution for AMR graphs as AMR only handle coreference resolution for entity and event variables within the sentence boundary. Interestingly, the researchers working on the abovementioned tasks also face the similar challenges and need to use primitive string matching rules as a temporal solution. Indeed, the traditional approaches to this problem, which heavily relies on formal features, cannot easily transferable to AMR graphs. This section, therefore, discusses one of the pioneer attempts to address this unsolved AMR-based discourse level coreference resolution, capitalizing on the power of Centering theory (Grosz et al., 1995) as well as the state-of-the-art rule-based coreference resolution approach (Lee et al., 2011).

3.3.1 Theoretical Framework

To begin with, Centering theory models the local component of attentional state which, in turn, is one of the three elements of discourse structure developed by Grosz and Sidner (1986) and, by definition, reflects "the discourse participant's focus of attention at any given point in the discourse" (Grosz et al., 1995). This local component is responsible for the local coherence of a discourse segment by reducing the inference load required to correctly interpret the focused entities – centers – which play the role of discourse connectors of adjacent utterances. Specifically, a single backward-looking center $C_b(U_i)$ and a partially ordered list of forward-looking centers $C_f(U_i)$ are usually defined for each utterance $U_i$ to represent its currently focused entity and the potential candidates for the focused entity in the next utterance, respectively. The trivial conception is that $C_b(U_i)$ belongs to $C_f(U_i)$, and the non-trivial one, also the key constraint of Centering theory, is that $C_b(U_i)$ is the highest-ranked entity of $C_f(U_{i-1})$ that is realized in $U_i$. With the assumption that
the task of $C_r$ ranking can be solved by the distribution of grammatical functions in an utterance (at least for English) (Brennan et al., 1987), the centering model is very compelling for pronominal anaphora resolution as one can use it to define the antecedent, the highest-ranking forward-looking entity in the previous utterance, of a backward-looking pronoun as well as other approximations.

To address the limited local scope of Centering theory, Walker (2000) introduces the cache model of attentional state that allows centering to interact with global discourse structure. This model is very compelling for coreference resolution in general and pronominal anaphora resolution in particular in the sense that, first, it does not restrict the scope of resolution as an antecedent candidate which is currently not in the cache can still be retrieved from the main memory; and second, it takes into consideration the discourse recency by identifying an adequate number of utterances as the cache size. It is worth noting that this recency is one important factor to determine the level of attentional focus or the salience of a discourse entity.

Turning back to the original formulation of Centering theory, we easily recognize that the salience of a discourse entity is materialized implicitly by its fixed local scope, i.e. its recency, and explicitly by its syntactic roles in the corresponding utterance. However, this fact should not mislead us about the strong semantic foundation of centering. In fact, Grosz et al. (1995) emphasize that “centers are semantic objects, not words, phrases, or syntactic forms.” In other words, the heavy reliance of current centering-based pronominal anaphora resolution on syntactic information is more or less for the sake of convenience rather than the optimal solution. Therefore, it is sensible to aim for more comprehensive models of salience that can robustly adapt to the deeper semantic representation layer of
discourse entities. Actually, one of such models can be inferred from Kaiser (2006), namely the multiple-factor model of salience, in which each factor is assigned an adequate weight to effectively reflect the overall salience score. In addition, the candidates for salience factors can be chosen from such dimensions as subjecthood, structural focus, pronominalization and givenness.

While Centering theory has been used successfully in pronominal anaphora resolution, it is not sufficient for the broader task of coreference resolution whose anaphors include not only pronouns but also proper names and definite descriptions. In fact, state-of-the-art coreference resolution systems typically exploit plentiful lexical, syntactic, semantic and discourse features in elaborate configurations. For example, Stanford’s multi-pass sieve coreference resolution system (Lee et al., 2011) is the collection of deterministic models (sieves) incorporating different string-based and semantics-based strategies and arranged in a pipeline manner in which the output of the model with the stronger features, i.e. higher precision, becomes the input of the model with the weaker features, i.e. lower precision. The system’s general preference order of feature application is: string-based → semantics-based → pronoun-based (including semantic agreements such as number, gender, person and animacy), which implies the plausibility and adequacy of the application of centering model at the end of the coreference resolution pipeline.

3.3.2. Task Description

To the best of our knowledge, there is currently no published work directly addressing the task of coreference resolution for AMR texts – coherent sequences of AMR graphs. The task itself, therefore, has yet been systematically formulated on the agenda of
the research community. As a result, this work first articulates a sensible task description that solidly captures the most relevant and universal aspects of the AMR data structure and content for diverse applications in future. It is also worth mentioning that, generally, there are two main categories of anaphoric relations namely identical and inferrable. The former is more popular and is the one addressed in this work; while the latter is usually associated with the more specific term – bridging anaphora resolution.

3.3.2.1. Working Corpus

There are two publicly available AMR-annotated corpora: the first one is *The Little Prince* Corpus, based on the English version of the novel *The Little Prince* by Antoine de Saint-Exupéry, and the second one is Bio AMR Corpus, consisting of cancer-related PubMed articles. The former is a better choice for this work at least in terms of the language universality and popularity. For the purpose of initial experiments, the first chapter of *The Little Prince* Corpus (version 1.5) is selected as the development set, containing contains 35 AMR (sentence-level) graphs, 308 AMR nodes and 311 AMR edges in total. For the evaluation, the second chapter of the same corpus is selected as the test set, containing 67 AMR (sentence-level) graphs, 557 AMR nodes and 538 AMR edges in total. In addition, the xml format of this corpus is very convenient for the integration of additional (light) annotation such as the information of cross-sentential coreference links between AMR nodes.
3.3.2.2. Anaphoric Candidates

The first step of coreference resolution is the identification of anaphoric candidates. In the tradition setting of the task, these candidates are usually called mentions (of entities) and typically realized in the form of noun phrases including pronominal, named, and definite nominal (Pradhan et al., 2011). However, the linguistic constituents of AMR graphs are concept nodes that are abstract away from the syntactic form, and, therefore, have different classification schemes. Based on their own semantic content and local semantic environment, these concept nodes can be divided into constant concepts and variable concepts that in turn consist of events (predicates with sense tags), special concepts (predefined concepts for labeling special phenomena such as wh-questions, ordinals, quantities, mathematical operators, dates, times, percentages, phones, emails and urls), logical conjunctions, deixis, named entities, (nominative) pronouns and others. They can also be grouped by the semantic roles they play in a graph including the core roles, corresponding to various indexed arguments of a predicate frame, and non-core roles, covering other semantic relations. These roles are encoded in the labels of the incoming edges of nodes, while their inverse counterparts (if available) are encoded in the labels of the outgoing ones. For example, in the graph shown in Figure 2, “swallow-01” is an event concept, while “boa” is a core-role concept as it is the agent argument of predicate “swallow-01”. Notice that ‘it’ is a controversial node that is not consistent with AMR Guidelines\(^{14}\) related to such phenomena as the main verb “be” and intra-sentential coreference; one can argue that this node should not exist as it semantically dissolve in “picture”.

\(^{14}\)https://github.com/amrisi/amr-guidelines/blob/master/amr.md
To decide which concept nodes should be selected as anaphoric candidates, we can first rely on certain intuition. Taking another look at Figure 2, we more or less agree that “picture”, “boa” (modified by “constrictor”) and “animal” are the most relevant anaphoric candidates. Indeed, this intuition goes along with researchers’ observations that “coreferent mentions are likely to appear as core verbal arguments” (Recasens et al., 2013) and, more abstractly, “nouns and entities typically encode much of the important semantic information in language” (Hill et al., 2015). Hence, following these conceptions, we should choose non-event concepts that are core-role or root nodes as the anaphoric candidates for this work and suggests the same approach for any applications of general purposes. It is worth noting that logical conjunction nodes are desirably treated as lists of their children.

**Figure 2:** AMR graph of a sentence in *The Little Prince* corpus.
concepts; while the aforementioned constants, special concepts, and deixis can be eliminated from the anaphoric candidates. In addition, as AMR does not represent quotation marks, it is non-trivial, for example, to define the true context of the first and second person pronouns (I, you, we, y'all); and therefore, coreference resolution for these pronouns in quotes is not taken into account in this work. Excluding the first sentence in the examined text, we can identify 100 anaphoric candidates, about one third of the total node number we have in the development set, and then detect 58 coreference links for these anaphoric candidates.

3.3.3. Algorithm Description

The proposed algorithm captures both spatial and temporal dimensions of the search for antecedents by using the cache model described below. It also features versatile and scalable configurations for the integration of various string-based and semantics-based features for different types of anaphoric candidates. Last but not least, it comprehensively models the concept of salience for ranking antecedent candidates.

3.3.3.1. Cache Model

The proposed algorithm handles one AMR graph at a time and defines the search space of antecedent candidates for anaphoric candidates in that graph as the (reversed) list of all previous (indexed) graphs in the corresponding AMR text. This search space is divided into cache and non-cache memory, which reflects the information recency. Specifically, the (changeable) cache size indicates the number of AMR graphs right before the AMR graph under consideration that form the cache memory. The rest of AMR graphs
appearing before cache area are the non-cache portion. As all of the AMR graphs are indexed and arranged in the reversed order, the algorithm essentially adds the temporal dimension into the search space, which allows the search to be performed in more recent graphs first. In example shown in Figure 3, the AMR graph under consideration is indexed \( i \), the cache (with cache size = 2) includes AMR graphs with indexes \( i-1 \) and \( i-2 \), and the AMR graphs whose indexes in the reversed range of \((i-3,1)\) belong to non-cache memory.

![Figure 3: Cache model.](image)

Figure 3 also shows us the advantage of the cache model in the sense that it allows three different types of search configuration: type 1 applies the same search strategy for both cache and non-cache memories; type 2 applies different search strategies for different memory portions; and type 3 restricts the search space to the cache, i.e. no search attempt is performed in non-cache memory. These different search types will be flexibly utilized for various coreference features described in the next section.
3.3.3.2. Coreference Features

As mentioned before, string-based and semantics-based features are commonly used in traditional rule-based coreference resolution systems. Specifically, string-based features are usually applied first to find antecedents for named and definite nominal anaphors; while semantics-based features are reserved for less apparent cases of definite nominal and pronominal anaphors. Following this approach, we can have the similar strategy for AMR anaphoric candidates by classifying them into three groups of named entities, (nominative) pronouns and ‘other’ concepts.

However, unlike the traditional approach in which one usually finds that string-based features are more reliable and diverse, the essence of AMR annotation scheme results in the opposite effect: string match between two nodes is very simple as each node is only a word in its base form; in contrast, the fact that each node is surrounded by well-defined semantic relations and concepts opens a new door for the exploration of semantics-based feature. Currently, the proposed algorithm uses only one string-based feature namely true-case full-string match (though allowing the integration of unlimited string-based features). Regarding semantics-based features, it implements a much richer plan, notwithstanding its limited use of external semantic resources. Basically, the implemented semantics-based features are relatively divided into intrinsic and extrinsic (to AMR content). The intrinsic features help in detecting first-time mentioned concepts, which cannot have an antecedent, and defining more specific senses of concepts under consideration. For example, indefiniteness, associated with first-time mentioned concept, is usually encoded by numerals or indefinite determiners and quantifiers such as another, any, some, many, much, a lot, few, little; therefore, all concepts modified by these
indefiniteness triggers should be removed from the list of anaphoric candidates. It is worth noting that articles are not included in AMR graphs and therefore cannot be exploited. In addition, AMR formalism handles nouns that invoke predicates by inserting general (semantically poor) concept nodes such as “thing” shown in Figure 4; therefore, instead of examining semantics of this kind of general concepts, the algorithm takes into account their child nodes, e.g. “copy-01” or “draw-01” in Figure 4.

Regarding the extrinsic semantic-based features, the algorithm mainly relies on WordNet to identify the gender and animacy properties of concepts as well as the (lowest common hypernym) semantic similarity of concept pairs and use them in different combinations to constrain the matching decision. In addition, one may notice that plurals are not represented in AMR graphs and therefore the number feature, which is frequently used traditional coreference resolution, cannot be implemented as well.
Figure 4: AMR graph including “thing” nodes.

Finally, it is worth noting that the algorithm implements the string-based match for non-pronominal concepts as well as the first and second person pronouns in the search space of type 1, the semantics-based match for non-pronominal concepts in the search space of type 2 (with different combinations of constraints for different memory types), and the combination of semantics-based match and salience-based ranking (see more detail in the next section) for the third person pronouns (she, he, it, they) in the search space of type 3. It is also necessary to mention that even though named entities and ‘other’
concepts are grouped in the non-pronominal category and generally treated in the same manner, the difference does exist; for example, if two nodes pass the string or semantic match and they both are named entities, they also need to pass the name match to be considered co-referred. Figure 5 shows a comprehensive example that demonstrates abovementioned strategies for coreference resolution.

![Diagram of coreference resolution](image)

**Figure 5**: Various strategies for coreference resolution.

### 3.3.3.3. Salience Ranking

While the string-based and semantics-based do help in identifying antecedent candidates, they are incapable of filtering the best ones, which results in poor coreference resolution, especially for the third person pronouns that only rely on the gender and
animacy constraints in the quest of their antecedents. To solve this problem, the algorithm calculates the salience ranking for relevant concepts and uses this ranking as preference for coreference resolution, essentially following the spirit of Centering theory.

As mentioned before, the concept of salience is perceived as a multi-factor model. The first factor of this model is the inverse value of shortest distance of a node from the root node; the shorter this distance is, the more focus this node receives. The second factor is the number of core roles a node plays that is weighted based on the index of each core role type; the smaller the index is, the more prominent the node is. Actually, this factor mimics the traditional syntactic ranking of Centering theory as there are strong correlations between the core roles and the subject-object functions (they all are the various arguments of the predicate). The salience model also includes two optional factors namely the pronominalization and the givenness: if a node is a pronoun or an anaphor, it is likely to be more salient. Further, each salience factor is assigned a weight, which is currently hand-crafted, to reflect its adequate contribution to the salience model. The final salience ranking score is the sum of these weighted factors.

Finally, it worth noting that the salience ranking is still partial, i.e. it is possible that there are more than one node have the same ranking. Therefore, in the case when the final resolution output consists of more than one antecedent candidate, the algorithm will accept all variants and assign an equal probabilistic weight for each coreference link such that the sum of assigned weights is equal to 1.
3.3.4. Experiments

3.3.4.1. Evaluation Metric

It is easy to recognize that the outputs of the proposed algorithm for AMR-based coreference resolution are coreference links. Thus, the link-based MUC metric (Vilain et al., 1995) is the most suitable option for the evaluation. However, as each output link of the proposed algorithm is associated with a probabilistic weight, we do not simply count the number of links but need to weight them before putting the weighted values into MUC formula.

3.3.4.2. Experiments on Development Set

We first tune and evaluate the proposed algorithm on the development set. Table 3, for example, shows the effect of each feature type (the cache size is fixed at 2).

<table>
<thead>
<tr>
<th>Features</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>String-based</td>
<td>61.9</td>
<td>22.4</td>
<td>32.9</td>
</tr>
<tr>
<td>+Semantics-based</td>
<td>75.0</td>
<td>36.2</td>
<td>48.8</td>
</tr>
<tr>
<td>+Pronoun-based</td>
<td>81.7</td>
<td>84.5</td>
<td>83.1</td>
</tr>
</tbody>
</table>

Table 3: Feature effects (cache size = 2).

As the Recall value is much lower than the Precision value in the first two result lines in Table 3, we can imply that both string-based and semantics-based features are effective but do not cover a large portion of coreference links for pronominal anaphors. The situation totally changes in the final result line in Table 3 when all features are taken into consideration: we achieve relatively balanced results with a slightly better Recall value in
comparison with the Precision value, which demonstrates the effectiveness and comprehensiveness of the designed feature combination.

Next, we try to vary the cache size to find the optimal one for this dataset. The results of this experiment are presented in Table 4.

<table>
<thead>
<tr>
<th>Cache size</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>83.1</td>
<td>84.5</td>
<td>83.8</td>
</tr>
<tr>
<td>2</td>
<td>81.7</td>
<td>84.5</td>
<td>83.1</td>
</tr>
<tr>
<td>3</td>
<td>80.3</td>
<td>84.5</td>
<td>82.4</td>
</tr>
</tbody>
</table>

*Table 4: Cache size effects.*

While the Recall values in Table 4 are the same, the Precision values slightly decrease when the cache sizes increase. The reasonable explanation for this phenomenon is that expanding the cache gives more space for false antecedent candidates to join the elite group and therefore introduces more guessing errors. These results are also in the favor of the original configuration of Centering theory in which only one previous utterance is taken into account for coreference resolution.

Finally, the contributions of optional salience factors to the salience ranking are illustrated in Table 5 (with cache size = 1, the best setting from the previous experiment).

<table>
<thead>
<tr>
<th>Salience factors</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>All factors</td>
<td>83.1</td>
<td>84.5</td>
<td>83.8</td>
</tr>
<tr>
<td>-Pronominal</td>
<td>83.1</td>
<td>84.5</td>
<td>83.8</td>
</tr>
<tr>
<td>-Given</td>
<td>88.1</td>
<td>89.7</td>
<td>88.9</td>
</tr>
</tbody>
</table>

*Table 5: Salience factor effects (cache size = 1).*

According to results in Table 5, the pronominalization factor has no effect on the effectiveness of the algorithm while the givenness factor really hurts the overall performance. However, it is still early to make any conclusion as these salience factors are
weighted by the hand-crafted values which may be suboptimal for this particular data set. In other words, we need more serious research on the task of finding the optimal weights for the multiple-factor model of salience presented in this work.

### 3.3.4.2. Final Result on Test Set

In the test set, we can identify 145 anaphoric candidates and annotate 75 coreference links for these anaphoric candidates. We use the cache size of one sentence, the optimal setting we get from the previous section, for the evaluation on this dataset. In addition, to obtain a better view on the controversy about the givenness factor's contribution to the salience model, we evaluate the proposed algorithm in both cases, when the givenness factor is present and absent respectively, as shown in Table 6. Interestingly, this result supports the positive contribution of the givenness factor to the overall performance, which goes against the finding of the experiments on the development set.

<table>
<thead>
<tr>
<th>Salience factors</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>All factors</td>
<td>75.3</td>
<td>68.1</td>
<td>71.5</td>
</tr>
<tr>
<td>-Given</td>
<td>71.3</td>
<td>66.1</td>
<td>68.6</td>
</tr>
</tbody>
</table>

Table 6: Evaluation result on the test set (cache size = 1).

### 3.4. Document-Level AMR-based Semantic Representation

This is the final step to build up a robust text-based semantic representation for ELL interaction that include two sub-steps: first defining the linking concepts between sentence-level AMR graphs of each text in Cinderella corpus by applying the algorithm of document level AMR-based coreference resolution developed in the previous section, and then merging these graphs into a bigger document-level AMR graph by using the defined
linking concepts and a series of graph operations such as graph union, vertex identification, and additional dummy root node creation.

### 3.4.1. Definition of Linking Concepts between Sentence-Level AMR Graphs

The definition of linking concepts between sentence-level AMR graphs of a text is essentially the coreference resolution for this text in the sense that each anaphoric concept node will be merged to it antecedent, which creates the physical cross-sentential connections for the examined text. Specifically, the developed algorithm and its best experimental configuration mentioned in the previous section are applied to all texts in Cinderella corpus. The single change in the implementation is the use of lower-case full-string match instead of true-case full-string match in terms of string-based features. The reason behind this decision is the observation that the proper names are not consistently capitalized in the output of CAMR parser. In other words, it makes the constraint of string match softer to capture this unavoidable shortcoming. Several AMR-based statistics of Cinderella corpus as the inputs and outputs of this coreference resolution process are presented in Table 7.

<table>
<thead>
<tr>
<th>No.</th>
<th>Text name</th>
<th>Node count</th>
<th>Edge count</th>
<th>Anaphoric candidate count</th>
<th>Coreference link count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ariel</td>
<td>315</td>
<td>277</td>
<td>140</td>
<td>74</td>
</tr>
<tr>
<td>2</td>
<td>Aurora</td>
<td>416</td>
<td>369</td>
<td>177</td>
<td>113</td>
</tr>
<tr>
<td>3</td>
<td>Belle</td>
<td>317</td>
<td>280</td>
<td>129</td>
<td>72</td>
</tr>
<tr>
<td>4</td>
<td>Cinderella</td>
<td>347</td>
<td>301</td>
<td>169</td>
<td>95</td>
</tr>
<tr>
<td>5</td>
<td>Snow White</td>
<td>258</td>
<td>215</td>
<td>95</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td><strong>Total:</strong></td>
<td><strong>1653</strong></td>
<td><strong>1442</strong></td>
<td><strong>710</strong></td>
<td><strong>403</strong></td>
</tr>
</tbody>
</table>

Table 7: AMR-based statistics of Cinderella corpus.
3.4.2. Merging of Sentence-Level AMR Graphs into Document-Level AMR Graph

Based on the original sentence-level AMR graphs and the linking concepts defined from previous step, the document-level AMR graph for a complete text is iteratively built up from the AMR graph of the first sentence in the text by merging with the AMR graph of the next sentence through two operations: graph union and vertex identification (or vertex contraction).

The first operation - graph union of two graphs - creates a new graph “whose vertex-set and edge-set are the disjoint unions, respectively, of the vertex-sets and the edge-sets of [the input graphs]” (Gross et al., 2013, p. 16). Visually, this operation put the second graph next to the first one on the same surface. It is worth noting that the input graphs of this operation must be disjoint, i.e. their vertex-sets must include different vertexes (or nodes). Therefore, all the nodes, or more exactly the variable name of these nodes, in each sentence-level AMR graph are relabeled by adding the corresponding sentence ID numbers as a prefix. For example, node $x$ in sentence $i$ will be relabeled as $i.x$. This additional information also beneficial for to semantic contents of the document-level AMR graph as a whole because it shows the temporal dimension of the text discourse which usually correlate with the order of the sentences in the text.

The second operation - vertex identification - is the generalization of edge contraction consisting of “the identification of [the edge’s] endpoints... (keeping the old adjacencies)” (Gross et al., 2013, p. 748). In other words, vertex identification is the merging of two vertexes which may be not connected to each other by any (incident) edge. Visually, this operation moves the second graph toward the first one by putting the identified nodes of the second graph on top of the corresponding nodes of the first one. In
case there are no identified nodes, a dummy root node is used to merge these graphs. In addition, the identified nodes in the result (document-level) AMR graph have a distinct attribute named *contraction* that contains the information of their co-referred nodes removed from the graph via vertex identification. This attribute turns out to be very useful for further information acquisition such as the occurrence frequency of a specific concept in the examined text.

As the AMR graphs in this project are developed from the data structure of MultiDiGraph of NetworkX\(^\text{15}\), we can capitalize on all available graph algorithms implemented by NetworkX that include all above-mentioned operations. Consequently, the merging algorithm is relatively straightforward as shown in Table 8.

<table>
<thead>
<tr>
<th>Input:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• TG - a list of sentence-level AMR graphs of a text</td>
</tr>
<tr>
<td>• LD - a list of sentence-level dictionaries whose keys are anaphoric nodes and whose values are the corresponding antecedent nodes (in previous sentences)</td>
</tr>
</tbody>
</table>

let \( T \) be an empty AMR graph

for (each sentence index \( i \) in \( TG \))

assign the graph union of \( T \) and \( TG[i] \) to \( T \)

if \( LD[i] \) is empty

perform vertex identification of dummy root nodes

else

for (each anaphoric node \( u \) in \( LD[i] \))

get antecedent node \( v \)

if \( v \) is in \( T \)'s nodes

perform vertex identification of \( v \) and \( u \) (\( u \) will be removed)

else

for (each node \( n \) in \( T \)'s identified nodes (having attribute *contraction*))

if \( v \) is in the *contraction* information of node \( n \)

perform vertex identification of \( n \) and \( u \) (\( u \) will be removed)

return \( T \)

Output: \( T \) - a document-level AMR graph of the examined text

---

\textbf{Table 8:} Merging algorithm for document-level AMR graph.

\(^\text{15}\)https://networkx.github.io/
The result document-level AMR graph of *Cinderella* story is visualized in Figure 6.

**Figure 6:** Document-level AMR graph of *Cinderella* story.
4. **Semaland Game - Proof-of-Concept Dialog System for ELL interaction**

It is recognizable that the document-level AMR graph for a text developed in the previous chapter satisfies the key requirement of a robust text-based semantic representation for ELL interaction discussed at the beginning of this work. Specifically, this developed semantic network for a whole text with multiple edges between pairs of semantic concept nodes allows the personalization of learners’ meaning-focused exploration of the text as they can travel through this multidimensional graph following their own path based on their specific semantic orientation. Consequently, the produced text-based interaction is capable of offering learners the genuine opportunities to promote their own development - self-scaffolding.

Moving forward, this chapter discusses the design of a proof-of-concept dialog system which provides robust meaning-focused interaction for ELLs based on the developed document-level AMR graph. Particularly, this discussion includes the description of *Semaland* game, the targeted unit of interactive communication between the dialog system’s conversational agent and ELLs, the development of the dialog system’s component, and the illustration of a real-life interaction example.

**4.1. Description of Semaland Game**

As communicative interaction is a discourse-level phenomenon, it is reasonable to apply studies in Formal Pragmatics as theoretical frameworks for its design. In particular, the idea of language game originating in Lewis’s (1979) seminal work is a sensible prompt
for this task. Rephrasing Lewis’s proposal, Roberts (2015) concisely defines language game as “a cooperative endeavor ... whose participants have common goals - roughly, the accurate sharing of information[, hence] are motivated to behave in an accommodating fashion.” Within this framework, Roberts (2012) describes conversational discourse “as a game, with context as a scoreboard organized around the question under discussion by the interlocutors.” This computational model is rationally selected as the skeleton of Semaland game, the interaction experience designed exclusively for the purpose of this work.

Specifically, Semaland game includes the following components:

- **Players**: the system’s conversational agent and the learner
- **Goals**: to explore semantic concepts and their relations in a text, having been read by both the agent and the learner, in a coherent manner
- **Rules**: the semantic concepts mentioned by the players must be relevant in the sense that they exist in the text (conventional rule) and their semantic relations in the text are close (conversational rule)
- **Moves**: the agent initiates the game with a text-level salient concept, the learner asserts another concept that satisfies the rules, then the agent and the user take turn to make the similar moves until there are $n$ concepts are listed where $n$ is an even number (to make sure the contributions from the agent and the learner are equal) and considerably smaller than the total number of concept nodes in the examined graph. The final pair of exchanges includes the agent’s request for the learner’s telling/writing a short story/essay demonstrating the semantic coherence of $n$ mentioned concepts and the learner’s respective response.
• Strategies: the learner should try to answer as quickly as possible to reflect their genuine impression of the text content, i.e. the information which is most prominently retained in his or her memory; meanwhile, the agent should choose the optimal concepts which have the closest relations with the previously mentioned concepts, especially the immediately previous one, in the course of the current game (to do that, the agent need to capitalize on the document-level AMR graph and certain graph-based distance calculations).

According to this description, the common ground of the players at the beginning of a game is the fact that both of them have read the text under consideration. After the initial utterance from the agent, the fact that the concept mentioned in this utterance is one of the most salient concepts in the text and the concept itself is added to the common ground. Similarly, for the rest of the turns taken by the user and the agent, relating to the concept assertion, the fact that the mentioned concepts are relevant according to the rules of the game and the concepts themselves are added to the common ground.

It is worth noting that the simplistic design of Semaland game is efficient in the sense that it greatly simplifies the later development of the proof-of-concept dialog system (see the discussion in the next section) yet convincingly demonstrates the desirable semantic interaction for ELLs because the learners, as emphasized before, are directly involved in the creation of the semantic dynamic of the interaction discourse. In addition, the special design of the final exchange of each interaction unit has a twofold purpose: to play the role of a comprehensive conclusion of a complete learning experience and to enhance learners’ retention rate of the text content by activating their productive skills, i.e. speaking and writing. The latter is especially relevant in the context of language learning.
which aims to the balance of learner’s competence (i.e. knowledge) and performance (i.e. communicative skills).

4.2. Implementation of *Semaland* Game Dialog System

In this section, the discussion focuses on the development of a basic text-based dialog system for *Semaland* game that includes three components: Natural Language Understanding (NLU), Dialog Manager (DM) and Natural Language Generation (NLG). Unlike spoken language dialog systems, this system does not handle speech input and output and therefore does not consists of Speech Recognition and Text-to-Speech Synthesis modules.

As mentioned before, the optimal design of *Semaland* game for the proof-of-concept purpose substantially simplifies this task. Indeed, the predefined concrete interaction contents, i.e. the semantic concepts in the form of words, allows us to unload complicated NLP handling of the system and learners’ utterances from NLU and NLG. Specifically, NLU is essentially the stemming/lemmatization of the input words of semantic concepts so that the system can recognize them in the content of AMR nodes; while NLG includes prescribed templates of utterances with certain slots for the system to fill in the semantic concepts relevant to a particular game discourse. Consequently, DM is the only component which requires serious efforts in its implementation. The overall architecture of the developed DM is presented in Figure 6.

As we can see, all elements of *Semaland* DM can be grouped into three layers namely *Utilities*, *Decisions* and *Discourse*. The easiest way to understand the DM’s working mechanism is putting *Discourse* in the center of consideration. This layer includes the historical record of all semantic concepts mentioned by the system agent and the learner in
the course of a specific game and grounded in the relationship with the corresponding text-based AMR graph.

**Figure 7:** *Semaland* Dialog Manager.

Specifically, after stemmed/lemmatized by NLU, an input concept from the learner goes through “identified concept” module to be mapped to a specific concept node in the text-based AMR graph before added to *Discourse*. If no mapping is recognized, this concept is still added to *Discourse* as an unknown concept from the system’s perspective, which is one of error handling steps in DM. Regarding the system’s output semantic concepts, we distinguish two different cases: the initial concept and the follow-up concepts in *Discourse*. The former is randomly chosen from the list of $k$ most salient concepts in the AMR text, generated by “top-$k$ concept” module (with the assumption that $k$ is considerably smaller than the total number of concept nodes in the examined graph). The latter is selected in a
much more elaborate manner by taking into account all concepts present in Discourse and their appearance order. First, the middle semantic concept nodes between the two most recent concept nodes in Discourse (if available) become the candidates for the next concept uttered by the agent. If there are no such nodes, the neighbors of these most recent nodes are chosen. In the worst case, “top-k concepts” that are different from those in Discourse are the back-up solution. Further, the semantic distances between the candidates for the next concept and Discourse concepts are calculated, using “semantic distance” module, which filter the most relevant candidates for the next concept in terms of Discourse semantic coherence. One of the most transparent ways to calculate the semantic distance is to sum the shortest path lengths in the undirected version of the text-based AMR graph between each candidate and all grounded concepts in Discourse, weighted by their respective recency. Finally, the next concept is randomly chosen from these elite candidates and sent to NLG.

4.3. Demonstration of Semaland Game Interaction

Figure 7 presents a full interaction unit of Semaland game between the conversational agent of the developed dialog system and an imaginary ELL who has read Cinderella story. Variable i is the number of exchange interactions between the agent and the user, while variable k is the parameter of “top-k concept” module. The lines beginning with the hash-tag symbol is not included in the real conversation, and presented here only for explanatory purpose. Specifically, these lines show the output of Decisions layer of abovementioned Semaland DM which are equivalent to different concept nodes in the text-based AMR graph of Cinderella story. The presented information of each concept node is a tuple consisting of the node’s variable name, representative concept type and
representative concept string. For example, tuple (‘0.x12’, ‘n’, ‘Cinderella’) corresponds to node ‘0.x12’, in which 0 is the sentence ID index, and displays the fact that this node is semantically represented as a name entity, i.e. ‘n’, whose content is ‘Cinderella’. Pay our attention to the sentence ID indexes of these nodes, we can realize that the temporal dimension is currently not taken into consideration by Decisions layer as these indexes seem to be random in terms of temporal order. This, therefore, indicates one of improvement areas for this work.

Figure 8: semaland game demonstration.
5. Conclusion and Future Work

This thesis research has presented an innovative and systematic approach to the development of ILTS capable of creating personalized and meaning-focus interaction for ELLs in a robust manner, and therefore raises a timely awareness of the necessity of this application type among other state-of-the-art language learning resources. From theoretical perspective, the most significant contribution of this work is the concise yet comprehensive argumentation for the work per se in the context of global SLE. From technical perspective, the most valuable discussion of this work is its pioneer attempt at solving the problem of discourse level coreference resolution for AMR texts as sequences of AMR graphs. Inspired by Centering theory and Stanford's rule-based coreference resolution system, the proposed algorithm features a simple cache model for searching antecedent candidates, a versatile and scalable framework for coreference feature integration, and a multiple-factor model of salience for ranking antecedent candidates. Last but not least, another unique characteristic of this work is its end-to-end product development, whose outcome is a proof-of-concept text-based dialog system handling Semaland game, an exclusive in-house design that allows ELLs to truly interact with the system's conversational agent regarding the semantic concepts in the text they have read, and therefore provides the learners with beneficial individualized learning experiences.

Being the first step in the exploration of the new-generation applications for worldwide ELLs, this research essentially poses several compelling tasks for future work. In
particular, regarding the discourse level coreference resolution for AMR texts, the fact that the proposed algorithm achieves a promising MUC F1 score of 71.5, testing on the second chapter of AMR-annotated *The Little Prince* Corpus, strongly encourages a more serious evaluation on a larger dataset and the use of this algorithm as a reasonable baseline for future research. In addition, handling of quoted information in AMR graphs and identification of the optimal weights for the factors in the developed salience model are also worth deeper investigation. Consequently, one AMR-related research topic can be proposed is how to optimize the current AMR annotation guidelines by taking into consideration the discourse aspect, which in turn can positively affect the quality of automatic AMR parsers. From perspective of DM for the interaction based on text-based AMR graph, the calculation of semantic distance between a concept node and the nodes in the discourse historical record requires more insights into graph theory for better computation models that reflect the semantic relevance of the concept node under consideration more accurately.

Finally, it is worth mentioning that while the work covered in this thesis satisfies the content requirements of the master's degree level, to develop it into advanced scientific knowledge, we need to make its components evaluable. For example, regarding the coreference resolution task for AMR texts, the effectiveness of the proposed algorithm should be evaluated based on a widely used test set and in comparison with certain state-of-the-art systems of the similar task, which is not address in this thesis. From educational perspective, serious experiments with ELLs should be conducted to validate the true effect of the designed semantic interaction on their learning process.
References


Kate, R. J., & Wong, Y. W. (2010). Semantic parsing: The task, the state of the art and the future [Tutorial slides]. In *Tutorials of the 48th Annual Meeting of the Association for*


