

Towards Ontology-Based Descriptions of Conversations with Qualitatively-Defined Concepts

Barbara Gendron-Audebert¹, Gaël Guibon^{1,2} and Mathieu D'Aquin¹

barbara.gendron@loria.fr

(1) LORIA, Université de Lorraine, CNRS

(2) LIPN, Université Sorbonne Paris Nord

19th TOTH International Conference - June 5, 2025

Conversational agents in 2025

Already used?



Ever
disappointed?

A solution: **generation control?**

Constrained Generation

Prompt

Model

Generated Content

What are the causes and effects of deforestation?



LLM

Deforestation happens from farming, logging, and building, leading to habitat loss, climate change, and soil damage.

Control "Answer like a pirate"



What are the causes and effects of deforestation?



LLM

Land be cleared for farms and treasure, causin' lost critters, hotter skies, and vanishin' soil!

Generation Control Through a Hybrid Approach

What is a relevant generation control?

- Formally defined
- Consistently applied
- Oriented towards human knowledge [Varshney et al., 2024]

→ Interaction with **structured knowledge**

What kind of structured knowledge?

- Formalize concepts and relationships
- Perform reasoning

→ **Conversation-dedicated ontology** [Gruber, 1993; Vickery, 1997]

*Enabling **conversational control** from logical definitions of conversation aspects defined in a **dedicated ontology***

Research question

How can qualitatively-defined conversational features be quantitatively described to enable control in LLM-based dialogue systems?

Contributions

- Formalize subjective conversation aspects in an ontology
- Use these aspects to control LLM outputs for **Proficiency-Level Control** (CEFR level)

Common European Framework of Reference for Languages (CEFR)

PROFICIENT USER	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
INDEPENDENT USER	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
BASIC USER	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Figure 1 – CEFR levels official definitions [Council of Europe, 2001]

Common European Framework of Reference for Languages (CEFR)

PROFICIENT USER	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations .
	C1	Can understand a wide range of demanding, longer texts , and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
INDEPENDENT USER	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
BASIC USER	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Figure 2 – CEFR levels official definitions contain **subjective** statements

Constrained Generation for Proficiency-Level Control

Prompt

Model

Generated Content

*What are the causes and
effects of deforestation?*
[ProficiencyLevel: A1]



*Deforestation is when
we cut too many trees.*

*What are the causes and
effects of deforestation?*
[ProficiencyLevel: C1]



*While deforestation is primarily
driven by agricultural expansion,
logging, and infrastructure
development, it leads to severe
environmental consequences...*

Constrained Generation for Proficiency-Level Control

Prompt

What are the causes and effects of deforestation?
[ProficiencyLevel: A1]

Model

LLM

Generated Content

Deforestation is when we cut too many trees.

What kind of CEFR Level Modeling?

- **Feature-based** [Gaillat et al., 2022] instead of deep learning based [Kerz et al., 2021; Schmalz and Brutti, 2021]
- Relies on balanced CEFR-annotated datasets
- Enables **formal ontology definitions**

→ **Decision tree** for interpretable CEFR classification

Our CEFR Level Modeling Setup

Expert-annotated **datasets**:

- CEFR-S sentences [Arase et al., 2022]
- CEFR-T texts [Nallapati et al., 2016]

44 linguistic **features** computed from SpaCy:

- Linguistic properties
- Readability metrics

Flesch-Kincaid Grade Level [Flesch, 1948]

$$\text{FKGL} = 0.39 \left(\frac{\text{Total Words}}{\text{Total Sentences}} \right) + 11.8 \left(\frac{\text{Total Syllables}}{\text{Total Words}} \right) - 15.59$$

SMOG Index [McLaughlin, 1969]

$$\text{SMOG} = 1.0430 \sqrt{\text{Number of Polysyllables} \times \left(\frac{30}{\text{Number of Sentences}} \right)} + 3.1291$$

CEFR Level Modeling Results

- We kept all features with **non-zero** feature importance.

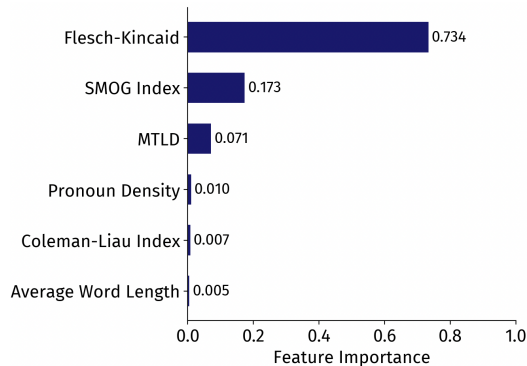


Figure 3 – Feature importances

- **MAE = 0.42**, Acc. = 0.66, wF1 = 0.65

Actual	A1	A2	B1	B2	C1	C2
	79	10	0	2	0	1
	23	41	4	2	0	1
	5	4	31	12	1	1
	1	2	15	45	18	2
	0	1	1	11	42	12
Predicted	A1	A2	B1	B2	C1	C2
	0	0	0	1	18	40

Figure 4 – Confusion matrix on CEFR-T test set

From Decision Tree Rules to Ontological Control

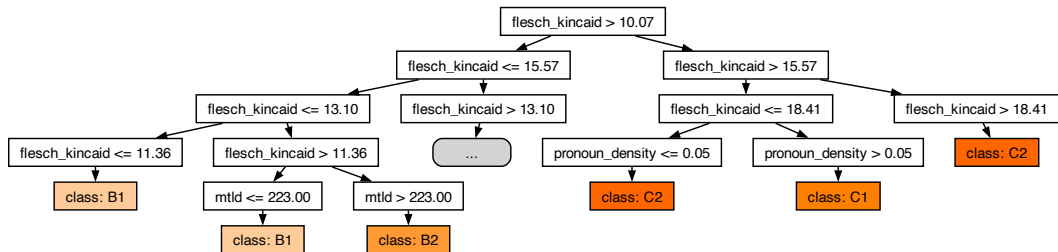


Figure 5 – Partial decision tree

Equivalent To 

- Utterance
and (hasCEFRLevel value B2)
- Utterance
and (((hasFleschKincaidReadabilityScore some xsd:decimal[> 11.358574390411377])
and (hasFleschKincaidReadabilityScore some xsd:decimal[<= 13.10405158996582])
and (hasMTLDMMeasure some xsd:decimal[> 223.0])) or ((hasFleschKincaidReadabilityScore some xsd:decimal[> 13.10405158996582])
and (hasFleschKincaidReadabilityScore some xsd:decimal[<= 14.32880687713623])
and (hasMTLDMMeasure some xsd:decimal[<= 370.25]))))

Figure 6 – Definition of B2LevelUtterance in the ontology

Proficiency-Level Control - Fine-Tuning Results

Evaluation method

- Input: random question + control prefix
- Evaluate through ontological reasoning on the output.

Model	F1 \pm Std	F1 Range	Acc	MAE
<i>Pre-Trained Baselines</i>				
Llama3-8B	0.06 \pm 0.10	0.00–0.29	0.16	2.42
Llama3.1-8B	0.14 \pm 0.07	0.09–0.30	0.19	1.98
Qwen2.5-7B	0.14 \pm 0.08	0.05–0.31	0.18	2.01
<i>Ours (Ontology-Guided Fine-Tuning)</i>				
Llama3-8B _F	0.31 \pm 0.01	0.15–0.44	0.19	1.22
Llama3.1-8B _F	0.22 \pm 0.05	0.17–0.29	0.23	1.57
Qwen2.5-7B _F	0.20 \pm 0.06	0.14–0.32	0.20	1.77

Table: Fine-tuning results for controlled generation over our CEFR level definitions

Proficiency-Level Control - Generation Examples

Prompt question: *What are the causes and effects of deforestation?*

Label	Generated Content
-------	-------------------

- | | |
|-----------|--|
| A1 | Forests are the lungs of the earth. Trees absorb carbon dioxide and emit oxygen. If we cut down too many trees, it could have a very bad effect on the global climate. |
| B2 | The main cause of deforestation is the conversion of land for agricultural use. Other significant causes are logging for the wood industry and wood pulp mills, as well as the expansion of cities. The effects of deforestation are far-reaching. |
| C2 | The main causes of deforestation are shifting agriculture, urbanization, and the establishment of plantations. The primary effects include global warming, water pollution, and the extinction of numerous species of animals and plants. Deforestation contributes to climate change because forests act as carbon sinks. |
-

*Enabling **conversational control** from logical definitions of conversation aspects defined in a **dedicated ontology***

- Quantitative-based formal definitions of subjective aspects
- Ontological control transferable to other use-cases

Perspectives

- Extension to more (complex) conversation aspects
- Conversation strategies → **merging descriptors**

An Example of Conversation Strategy

Job interview practice: proficiency level, emotional load (L), polarity (P).

Exit small talk

User: *Sure. [...] I used to be a data science intern and now I apply to my first full time position as a data scientist.*

Agent: *That's great! [...] Can you tell me more about your experience as a data science intern? What kind of projects did you work on during that time, and what skills did you develop? ($L = 0$, $P = 2$)*

From background to technical point

User: *I have real experience and I know how to adapt quickly to many different contexts and technical stacks. Also, if your company has a home-made framework, I can learn it quite fast.*

Agent: *Let's dive into the technical aspects of adapting to different frameworks and stacks. [...] How do you handle debugging and optimizing the performance of a cross-platform desktop application built with Electron [...] ? ($L = 0$)*

References I



Arase, Yuki, Satoru Uchida, and Tomoyuki Kajiwara. “CEFR-Based Sentence Difficulty Annotation and Assessment”. In: *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Ed. by Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, Dec. 2022, pp. 6206–6219.



Flesch, Rudolph. “A new readability yardstick.”. In: *Journal of applied psychology* 32.3 (1948), p. 221.



Gaillat, Thomas et al. “Predicting CEFR levels in learners of English: The use of microsystem criterial features in a machine learning approach”. In: *ReCALL* 34.2 (2022), 130–146. DOI: 10.1017/S095834402100029X.



Geertzen, Jeroen, Dora Alexopoulou, and Anna Korhonen. “Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database (EFCAMDAT)”. In: Jan. 2013.



Kerz, Elma et al. “Automated Classification of Written Proficiency Levels on the CEFR-Scale through Complexity Contours and RNNs”. In: *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications*. Ed. by Jill Burstein et al. Online: Association for Computational Linguistics, Apr. 2021, pp. 199–209. URL: <https://aclanthology.org/2021.bea-1.21>.



McLaughlin, G. Harry. “SMOG Grading: A New Readability Formula”. In: *Journal of Reading* 12.8 (1969), pp. 639–646.

References II



Nallapati, Ramesh et al. “Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond”. In: *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*. Ed. by Stefan Riezler and Yoav Goldberg. Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 280–290. DOI: 10.18653/v1/K16-1028. URL: <https://aclanthology.org/K16-1028>.



Schmalz, Veronica Juliana and Alessio Brutti. “Automatic Assessment of English CEFR Levels Using BERT Embeddings”. In: *Italian Conference on Computational Linguistics*. 2021. URL: <https://api.semanticscholar.org/CorpusID:245125483>.



Sun, Jiao et al. “Evaluating Large Language Models on Controlled Generation Tasks”. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Ed. by Houda Bouamor, Juan Pino, and Kalika Bali. Singapore: Association for Computational Linguistics, Dec. 2023, pp. 3155–3168. DOI: 10.18653/v1/2023.emnlp-main.190. URL: <https://aclanthology.org/2023.emnlp-main.190>.



Varshney, Deeksha, Asif Ekbal, and Erik Cambria. “Emotion-and-knowledge grounded response generation in an open-domain dialogue setting”. In: *Knowledge-Based Systems* 284 (2024), p. 111173. ISSN: 0950-7051. DOI: <https://doi.org/10.1016/j.knosys.2023.111173>. URL: <https://www.sciencedirect.com/science/article/pii/S0950705123009231>.