Introduction to Machine Translation

Machine Translation: Advanced Topics

Vincent Vandeghinste

KU Leuven – Brussels

2026

Outline

What is Machine Translation?

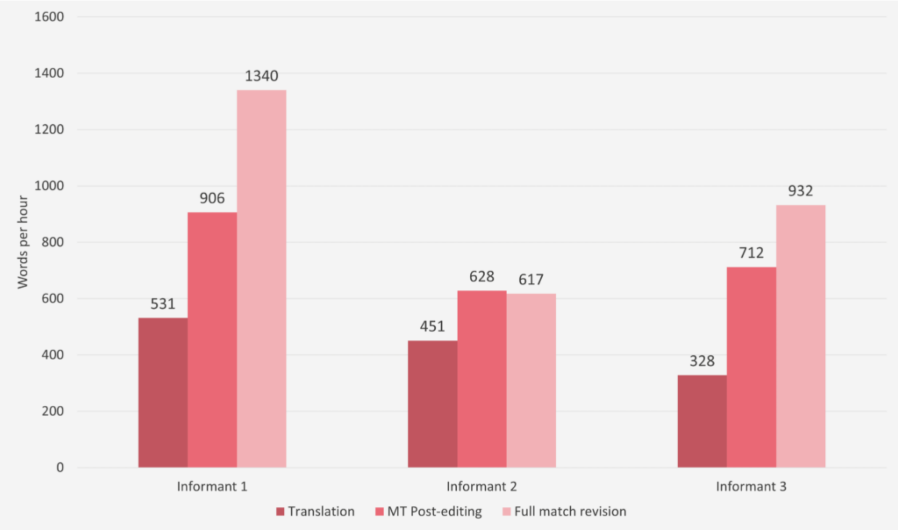
What is Machine Translation?

- Machine Translation (MT) is commonly defined as *the use of computers to translate text from one natural language into another*. (European Association for Machine Translation 2024)
- MT is one of the oldest ambitions in computer science.
- Fully automatic, high-quality translation remains an unsolved problem.
- Current MT systems can still produce useful output for many domains and tasks.

MT in practice

- Embedded in web applications
- Integrated in CAT tools
- Used for multilingual information retrieval
- Used for speech-to-speech translation
- Neural MT can substantially increase translator productivity (Plitt and Masselot 2010; CrossLang 2025)

Figure: Productivity increase with MT



Goal of the course Machine Translation: Advanced Topics

Goal of the course: Machine Translation: Advanced Topics

1. **Machine Translation** (these notes)
   - Introduces the MT research field and the current state-of-the-art
   - Hands-on experience in the experimental paradigm in MT research
   - Study effects of major breakthroughs in neural MT since ~2013
2. **Post-editing** (out of scope here)
   - Critical, hands-on study of post-editing processes

Machine Translation in the Context of NLP and AI

MT in the Context of NLP and AI

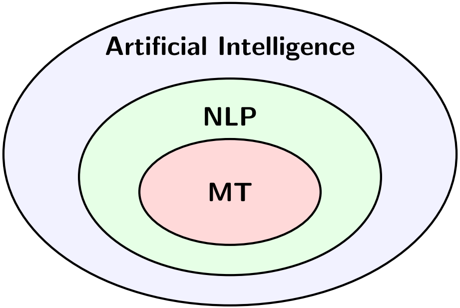


Figure: MT as a subfield of Natural Language Processing and Artificial Intelligence.

Artificial Intelligence (AI)

- Computational systems performing tasks associated with human intelligence
- Reasoning, decision making, perception, planning, learning, language *understanding*
- For MT: provides foundations for models that learn patterns from data and make predictions
- Modern MT is mainly part of *machine learning*

Natural Language Processing (NLP)

- Enables computers to analyse, generate, and interact using human language
- Tasks: PoS tagging, parsing, sentiment analysis, QA, IE, MT, ...
- Natural language is ambiguous and context-dependent
- NLP combines linguistic structure with statistical/neural uncertainty handling
- Interdisciplinary field (Jurafsky and Martin 2025)

Figure: NLP as an interdisciplinary field

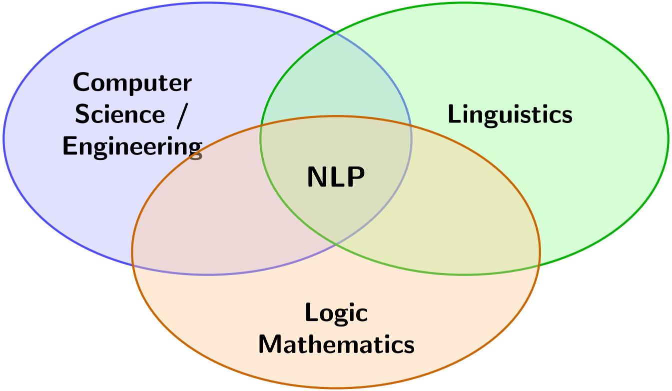


Figure: NLP as an interdisciplinary field.

Interdisciplinarity

- **Computer science:** MT is treated as an engineering problem, focusing on algorithms that automatically map an input string in one language to an output string in another language, with emphasis on efficiency, scalability, and measurable performance.
- **Linguistics:** MT is viewed as a problem of analysing and generating natural language, drawing on linguistic theories of morphology, syntax, semantics, and ambiguity to model how meaning is expressed across languages.
- **Logic:** MT is conceived as a meaning-preservation task, where translation operates over abstract or formal representations and aims to preserve truth conditions and inferential relations across languages.

Example: Linguistic ambiguity in NLP/MT

Illustrative ambiguity

I made her duck.

- Multiple interpretations involving lexical, syntactic, and semantic ambiguity
- MT systems must manage ambiguity via context, training data, and probabilistic scoring (Koehn 2020; Sennrich 2018)

Illustrative ambiguity: all readings I

I made her duck.

This short sentence has several distinct readings:

1. **Culinary reading (create + dish)**
   "I cooked a duck for her."
   make = 'prepare (food)', *duck* = noun, *her* = indirect object.
2. **Culinary reading (create + possession)**
   "I cooked the duck that belongs to her."
   make = 'prepare', *duck* = noun, *her* = possessive determiner.
3. **Causative reading (cause to move)**
   "I caused her to lower her head (to duck)."
   make = causative verb, *duck* = verb, *her* = direct object.

Illustrative ambiguity: all readings II

4. **Transformative reading (turn into a duck)**
   "I turned her into a duck."
   make = 'cause to become', *duck* = noun (new state).
5. **Magical creative reading**
   "I created the duck which belongs to her."
   make = 'create', *duck* = noun, *her* = possessive determiner.

Key linguistic challenges

- **Word sense disambiguation**: selecting the intended meaning of a polysemous word from context.
- **Anaphora resolution**: identifying which earlier expression a pronoun/referring element points to.
- **Structural divergence**: systematic differences in how languages encode the same meaning via syntax, morphology, or word order.

Professional Associations and Regional Conferences

The MT research community is structured around three major regional associations:

- **EAMT** (Europe), **AMTA** (Americas), and **AAMT** (Asia-Pacific) each support MT research and practice within their region.
- Each association organises a regional conference serving as the main MT forum for that region.
- Together, they form the **International Association for Machine Translation (IAMT)**.
- **MT Summit**, organised by IAMT, is the flagship international MT conference and rotates across regions.

Major International Conferences and Shared Tasks

Beyond regional conferences, several international venues play a key role in shaping MT research:

- **WMT** is the main global MT benchmarking venue, best known for its large-scale shared tasks covering news translation, low-resource MT, evaluation metrics, and quality estimation.
- **IWSLT** focuses on spoken and multimodal translation, providing benchmarks for speech-to-text, speech-to-speech, and simultaneous translation.
- **ACL** conferences regularly include a substantial body of MT research within the broader NLP community.
- **LREC** emphasises language resources and evaluation, with strong attention to datasets and smaller or under-resourced languages.
- **Shared tasks** provide standard datasets and evaluation protocols, enabling fair comparison, reproducibility, and progress tracking in MT research.

Shared Tasks in Machine Translation

Shared tasks are community-driven evaluation campaigns in which multiple research groups address the same MT problem under controlled conditions:

- common training, development, and test datasets are provided;
- evaluation metrics and protocols are fixed in advance;
- systems are compared in a fair and reproducible manner.
- results are presented and analysed in a coordinated workshop.

Shared tasks play a central role in MT research by establishing benchmarks, highlighting open challenges, encouraging methodological rigor, and creating reference points for future system development.

Open Access and Knowledge Dissemination

The MT research community is strongly committed to open scientific communication:

- The **ACL Anthology** provides open access to peer-reviewed research in NLP and Machine Translation.
- Most major MT venues publish their proceedings through the Anthology, ensuring long-term preservation and broad accessibility.
- Open access supports transparency in methods, datasets, and evaluation results.

Together with regional associations and shared tasks, open dissemination has enabled rapid scientific progress and industrial innovation in MT.

MT History: From Rule-Based Systems to Neural and LLMs

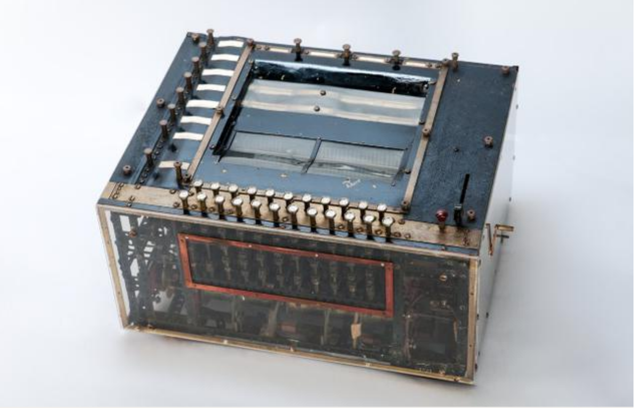
MT History: why paradigms matter

- Organised around paradigms that emerged to address limitations of earlier ones
- We introduce each approach when it historically emerged

Pre-digital concepts and early visions

- Universal language / interlingual ideas (Descartes; Wilkins)
- 1930s: mechanical aids for dictionary lookup (Artsrouni)
- 1949: translation as decoding problem (Weaver 1949)

Figure: Artsrouni's mechanical brain



Machine à traduire électromécanique dite le Cerveau mécanique, 1933-1935. Inv. 21661.

Figure: Le Cerveau Méchanique.

1950s–1970s: Rule-Based Machine Translation (RBMT)

- 1954 Georgetown–IBM experiment (Hutchins 2004)
- Video about early MT
- Translation via explicit linguistic knowledge:
  - morphological analysers, grammars, bilingual lexicons, transfer rules
- Vauquois triangle: direct → transfer → interlingua (Vauquois 1968)

Figure: Vauquois triangle

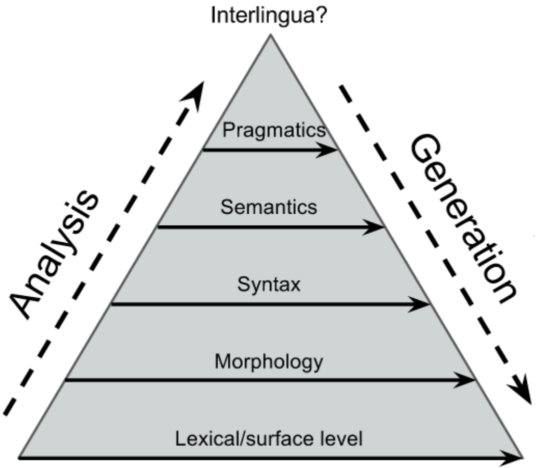


Figure: The Vauquois triangle, illustrating translation strategies.

Direct translation (surface level)

- Operates on lexical/surface level
- No explicit syntactic structure
- Assumes similar agreement and word order

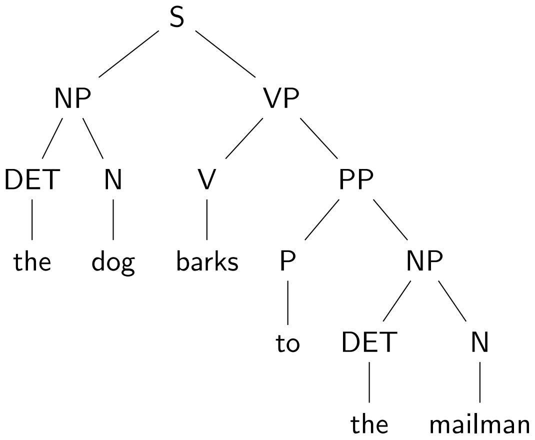
Direct Translation Example

Step	Operation
Input	`she eats apples`
Morphological analysis	`eats = eat + 3rd person singular`
	`apples = apple + plural`
Lexical lookup	`she → zij`
	`eat → eten`
	`apple → appel`
Morphological generation	`eten + 3rd person singular = eet`
	`appel + plural = apples`
Output	`zij eet appels`

Transfer-based translation: analysis

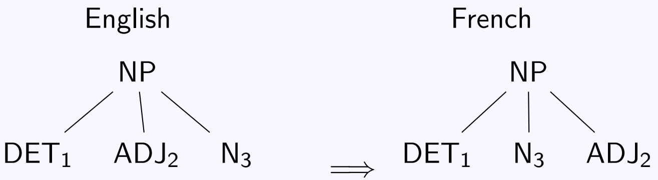
- Source sentence is syntactically (and possibly semantically) parsed
- Output: a structured representation (e.g., a tree)

Example Syntactic Analysis Sentence: `the dog barks to the mailman`



Syntactic transfer rules (STSG)

Example transfer rule between English and French

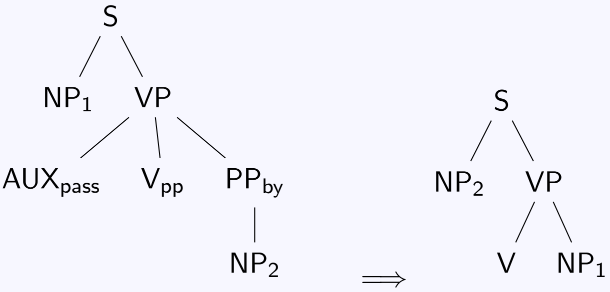


Example:

$\texttt{DET}_1\texttt{[the]}$ $\texttt{ADJ}_2\texttt{[blue]}$ $\texttt{N}_3\texttt{[sky]}$ $\rightarrow$ $\texttt{DET}_1\texttt{[le]}$ $\texttt{N}_3\texttt{[ciel]}$ $\texttt{ADJ}_2\texttt{[bleu]}$

Transfer example: Passive ↔ active

A more complex transfer example: Passive ↔ active



Example:

$\mathtt{NP}_1$[de man] $\mathtt{AUX}_{pass}$[wordt] $\mathtt{V}_{pp}$[gebeten] $\mathtt{PP}_{by}$[door $\mathtt{NP}_2$[de hond]] $\rightarrow$
$\mathtt{NP}_2$[the dog] V[bites] $\mathtt{NP}_1$[the man]

Semantic transfer

- Transfer rules operate on predicate–argument structures
- Abstract away from word order
- Robust handling of alternations (passive, argument reordering)

Semantic Transfer Rules

Predicate mapping: `eat(x,y) → eten(x,y)`

Argument Reversal: `miss(x,y) → manquer(y,x)`

Role preservation: `AGENT(eat,x) → AGENT(eten,x)`
`PATIENT(eat,y) → PATIENT(eten,y)`

Semantic transfer: example

Source Analysis:
```
eat(
  AGENT = she,
  PATIENT = apples,
  TENSE = present
)
```

Transfer rules:
```
eat(x,y)  → eten(x,y)
she       → zij
apples    → appel[PL]
```

Target generation:
```
S → NP V NP
NP(AGENT) = zij
V = eet
NP(PATIENT)[PL] = appels
```

Interlingua-based translation (deep semantic level)

- **Interlingua representation**:

  ```
  EVENT: EAT
    AGENT: FEMALE_PERSON
    PATIENT: APPLE (PLURAL)
    TENSE: PRESENT
  ```

- **Source analysis**: `she eats apples`
  $\Rightarrow$ `EAT(AGENT=FEMALE_PERSON, PATIENT=APPLE[PL], TENSE=PRESENT)`
- **Target generation (Dutch)**:

  ```
  NP(FEMALE_PERSON) → zij
  V(EAT, PRESENT, 3SG) → eet
  NP(APPLE, PLURAL) → appels
  ```

- Maximises reuse across languages, but requires very rich semantic representations

Typical RBMT components

- Morphological analysis and generation (inflection, agreement, derivation)
- Syntactic analysis (phrase-structure or dependency grammars)
- Lexical transfer (often with PoS/sense information)
- Structural transfer rules (word order, argument structure, realisation)

RBMT: strengths and limitations

Strengths

- Precise and interpretable
- Strong in controlled domains

Limitations

- Immense manual effort
- Brittle to unexpected input
- Poor scalability across domains/language pairs
- Pipeline **error percolation**

- Interlingua has conceptual scalability (N vs. N^2 modules), but is hard in practice due to semantic-representation difficulty.

1980s: Example-Based Machine Translation (EBMT)

- Proposed by Nagao (1984): *translation by analogy*
- First clearly **data-driven** MT approach
- Basic operations:
  1. retrieve similar examples
  2. identify correspondences
  3. recombine translated fragments

EBMT: full example

Assume the parallel corpus contains:

Example 1:
EN: *I like green apples.*
NL: *Ik hou van groene appels*
Example 2:
EN: *I like red pears.*
NL: *Ik hou van rode peren*

To translate: `I like green pears.`

- match `I like` ↔ `Ik hou van`
- take `green` from Example 1
- take `pears` from Example 2
- recombine: `Ik hou van groene peren`

EBMT limitations

- Depends heavily on corpus coverage and diversity
- Reliable correspondences are hard (heuristic alignment)
- Limited generalisation beyond observed patterns
- Brittle with lexical variation, reordering, longer sentences

1990s–2000s: Statistical Machine Translation (SMT)

- Shift enabled by parallel corpora + compute
- Replace rules with probability distributions learned from data (Brown et al. 1993)
- Log-linear framework; approximate decoding
- Phrase-based SMT popularised via Moses (Koehn et al. 2007)

SMT core components

- Word alignment (IBM models etc.)
- Phrase-based translation (phrase tables)
- Target-side language modelling (n-grams)
- Log-linear scoring + decoding (beam search over hypotheses)

SMT example: phrase table + decoding steps I

For the sentence: `She eats apples.`
A simplified **phrase table** might contain:

Source phrase	Target phrase	Probability
`she`	`zij`	0.92
`eats`	`eet`	0.87
`apples`	`appels`	0.94
`apples`	`de appels`	0.04
`eats apples`	`eet appels`	0.76

During **decoding**, the SMT system incrementally constructs partial translation hypotheses and maintains a beam of the highest-scoring candidates.

Step 0: Initial state

- Covered source words: none

SMT example: phrase table + decoding steps II

- Hypothesis: empty
- Score: 0.0

Step 1: Translating `she`

- Apply: `she` → `zij`

Partial translation	Covered source	Score (log)
`zij`	`she`	-0.8

The score reflects the translation probability and an initial language-model contribution.

Step 2: Expanding with `eats`

- Apply: `eats` → `eet`

SMT example: phrase table + decoding steps III

Partial translation	Covered source	Score (log)
`zij eet`	`she eats`	-1.2
`zij`	`she`	-2.0

`zij eet` is preferred: more coverage + more fluent sequence.

Step 3: Translating `apples`

- `apples` $\rightarrow$ `appels`
- `apples` $\rightarrow$ `de appels`
- `eats apples` $\rightarrow$ `eet appels`

SMT example: phrase table + decoding steps IV

Partial translation	Covered source	Score (log)
`zij eet appels`	`she eats apples`	-1.5
`zij eet de appels`	`she eats apples`	-2.3
`zij eet`	`she eats`	-2.6

After all source words are covered, the highest-scoring hypothesis is selected.

SMT limitations (motivation for NMT)

- Reliance on local phrase translations and *n*-gram LMs
- Limited long-distance dependencies and global coherence
- Motivated transition to neural MT

2010s: Neural Machine Translation

- Neural MT emerged from:
  - neural language modelling (Bengio et al. 2003)
  - sequence-to-sequence learning (Sutskever et al. 2014)
  - attention (Bahdanau, Cho, and Bengio 2015)
  - Transformer architecture (Vaswani et al. 2017)
- Subword segmentation, back-translation, multilingual training extended NMT

2020s–present: Large multilingual models and LLMs

- Large multilingual encoder–decoders (mBART, mT5, NLLB)
- Decoder-only LLMs (GPT-style, LLaMA, Qwen)
- Enable zero-shot/few-shot translation and multilingual transfer
- Challenges: controllability, consistency, evaluation

The Experimental Paradigm in Machine Translation

Experimental paradigm in MT

- Goal: test whether a **specific design choice** improves quality, efficiency, or robustness
- Use controlled comparisons to support causal claims
- Keep all non-target variables constant

Minimal pairs of experimental conditions

- Compare systems differing in exactly one controlled variable
- Example manipulations:
  - number of layers
  - tokenisation (word vs subword)
  - training data amount/type
- Enforce discipline: fixed seeds, identical preprocessing, identical training budget

Controlled data splits

- Training: learn parameters
- Validation: tuning and early stopping
- Test: final evaluation only

Key principle

The test set is used *once*, after all design decisions have been made.

- Use shared-task predefined splits when available for comparability

Evaluation and significance testing

- Metrics: BLEU, chrF, COMET, BLEURT, ...
- Report multiple metrics (different aspects of quality)
- Test statistical significance (e.g., bootstrap resampling)
- Avoid overinterpreting small differences without testing

Interpretation and reporting

- Inspect example outputs; identify systematic error patterns
- Report training cost, inference speed, robustness effects
- Ensure reproducibility: data, preprocessing, architecture, hyperparameters, evaluation

The State of MT Today

The state of MT today

- Rapid technological change; expanding multilingual coverage
- Deep integration into professional translation workflows
- MT supports and reshapes CAT environments (Intento 2025)

Coexistence of NMT and LLM-based MT

- NMT remains production backbone (quality, predictability, low latency)
- LLMs add robustness to noisy input and better instruction/context use
- Hybrid setups: NMT base translation + LLM post-editing/style/terminology/QE

Integration in professional workflows

- MT in CAT tools, real-time suggestions
- Document-level processing for consistency and terminology control
- Customisation: glossaries, domain adaptation, interactive prompting
- Quality estimation to prioritise post-editing effort

Evaluation trends

- Beyond BLEU/chrF: document-level coherence, consistency, discourse
- Context- and task-aware evaluation; risk in sensitive domains
- Reference-free metrics and LLM-based evaluators under active study
- Human evaluation remains essential in many settings

Challenges and risks

- Controllability (terminology, style guides, constraints)
- Reliability (omissions, hallucinations)
- Uneven coverage (low-resource languages)
- Data governance: privacy, copyright, transparency
- Bias and fairness

References I

- Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio (2015). "Neural Machine Translation by Jointly Learning to Align and Translate". In: *ICLR*.
- Bengio, Yoshua et al. (2003). "A Neural Probabilistic Language Model". In: *Journal of Machine Learning Research* 3.
- Brown, Peter F. et al. (1993). "The Mathematics of Statistical Machine Translation: Parameter Estimation". In: *Computational Linguistics* 19.2.
- CrossLang (2025). *Human Evaluation of Machine Translation Quality – A Quick Guide/*.
- European Association for Machine Translation (2024). *What is Machine Translation?* URL: https://eamt.org/what-is-machine-translation/.
- Hutchins, W. John (Sept. 2004). "The Georgetown-IBM experiment demonstrated in January 1954". In: *Proceedings of the 6th Conference of the Association for Machine Translation in the Americas: Technical Papers*. Ed. by Robert E. Frederking and Kathryn B. Taylor. Washington, USA: Springer, pp. 102–114. URL: https://aclanthology.org/2004.amta-papers.12/.

References II

- Intento (2025). *State of Machine Translation 2025*.
- Jurafsky, Daniel and James H. Martin (2025). *Speech and Language Processing*. 3rd ed. Draft of August 24, 2025. Prentice Hall.
- Koehn, Philipp (2020). *Neural Machine Translation*. Cambridge University Press.
- Koehn, Philipp et al. (June 2007). "Moses: Open Source Toolkit for Statistical Machine Translation". In: *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*. Ed. by Sophia Ananiadou. Prague, Czech Republic: Association for Computational Linguistics, pp. 177–180. URL: https://aclanthology.org/P07-2045/.
- Nagao, Makoto (1984). "A framework of a mechanical translation between Japanese and English by analogy principle". In: *Proc. of the International NATO Symposium on Artificial and Human Intelligence*. Lyon, France: Elsevier North-Holland, Inc., pp. 173–180. ISBN: 0444865454.

References III

- Plitt, Mirko and François Masselot (2010). "A productivity test of statistical machine translation post-editing in a typical localization context". In: *Prague Bulletin of Mathematical Linguistics* 93, pp. 7–16.
- Sennrich, Rico (2018). "Machine Translation: Lecture Notes". In: Available online.
- Sutskever, Ilya et al. (2014). "Sequence to Sequence Learning with Neural Networks". In: *NIPS*.
- Vaswani, Ashish et al. (2017). "Attention Is All You Need". In: *NIPS*.
- Vauquois, Bernard (1968). "Structures profondes et traduction automatique. Le système du C.E.T.A.". In: *Revue roumaine de linguistique* XIII (2), p. 105. ISSN: 0035-3957.
- Weaver, Warren (1949). *Translation*. Memorandum, reprinted in Locke and Booth (1955).