# CHRISTOPHE SERVAN – NLP RESEARCH SCIENTIST AND MANAGER

## Personnal

| | |
|---|---|
| Firstname | Christophe |
| Lastname | Servan |
| Birth | $1^{er}$ mai 1979 in Cavaillon (Vaucluse – France) |
| Nationality | French |
| Family status | Married, 2 children |
| Personnal | 9 Chemin de la Landerie<br>F-78125, HERMERAY – France<br>+33 (0) 6 15 11 75 82<br>christophe.servan@gmail.com |
| Website | http://cservan.github.io/ |
| Research topics | |

- Natural & Spoken Language Understanding
- Information Retrieval
- Conversational Question–Answering
- Large Language Model scaling & ecological impact

## Work experience

**01/2022–present**   **Researcher**, *LISN–CNRS*, Orsay, France.
Researcher put–at–disposal: 80% au LISN-CNRS & 20% Research Manager at Qwant: Multilingual Spoken language understanding for Contextual Question/Answering (MuSCQA) – LLM & Contextual Conversational Search; management : 3 PhD students co-supervised & 3 internships (full supervision).

**01/2018–present**   **NLP Research Scientist and Manager**, *Qwant*, Paris, France.
Deep Learning & Artificial Intelligence for NLP (mainly NLU & MT) applied to Information Retrieval – NLP team Manager, leading scientific and founded projects, act as deputy chief science officer (CSO) answering directly to the CTO since 2019.

**09/2019–12/2022**   **NLP Professor**, *EPITA*, Paris, France.
In charge of NLP teachings in the International Program at EPITA – Part of the Train The Teacher program for the benefit of the Ministry of Communications and Information Technology of Egypt

**09/2016–01/2018**   **Applied Research Scientist**, *SYSTRAN (R&D)*, Paris, France.
Investigations on domain adaptation and the adding of lexical data in Neuronal Machine Translation – Information Extraction from tweets

**03/2015–08/2016**   **Post-doctoral Researcher**, *LIG (GETALP) – University of Grenoble Alpes*, Grenoble, France.
Investigations on Deep Learning for confidence measures & speech translation evaluation within a human or an automatic post-editing framework

**09/2013–02/2015**   **Research Scientist**, *Xerox Research Centre Europe (MLDAT)*, Meylan, France.
Enhancement of domain adaptation for the translation of post-edited speeches

**11/2010–08/2013**   **Post-doctoral Researcher**, *LIUM (LST) – University of Le Mans*, Le Mans, France.
Studying and Evaluating machine learning for adaptation and optimization of SMT, applied to text and speech, within a human post-editing framework

**10/2009–10/2010**   **Research Engineer**, *CEA–LIST (LVIC)*, Fontenay-aux-Roses, France.
Exploration of hybrid approaches for translation (speech and text) applied to cross-lingual Information Retrieval

**09/2008–09/2009**   **Temporary Assitant Professor**, *LIA (team Human–Machine Dialogue) – Avignon University*, Avignon, France.
Multilingual Spoken language Understanding & Human–Machine Dialogue

**09/2005–08/2008**   **PhD**, *LIA (team Human–Machine Dialogue) – Avignon University*, Avignon, France.
Machine Learning, Spoken language Understanding & Human–Machine Dialogue

## Degrees

| | |
|---|---|
| 09/2005–08/2008 | **PhD**, *LIA (Team "Human–Machine Dialogue ")*, Avignon University, defense : 10 december 2008, mention très honorable. |
| Advisors | Rénato De Mori et Frédéric Béchet |
| Jury | Sophie Rosset, Kamel Smaïli, Marc El-Bèze, Géraldine Damnati and Yannick Estève |
| Title | Machine Learning and Spoken Language Understanding within the framework of a mixed interactive human–machine dialogue by phone |
| 2004–2005 | **Research Master in Computer Science**, *topic : NLP and speech processing*, CERI – Avignon University, mention bien. |

## Management

| | |
|---|---|
| PhD | Co-supervised 1 PhD Student with the LISN, at Paris-Orsay University (2019–2022) |
| | Co-supervising 2 PhD Students with the LISN, at Paris-Orsay University (2022–2024) |
| Internships | Supervised 9 Master students at Grenoble University, QWANT and LISN (2016–2023) |
| Team | Managed a team of up to 5 people at Qwant, Topic lead (Conversational Search) LISN |

## Computer Science

| | |
|---|---|
| GitHub | https://github.com/cservan/ |
| HuggingFace | https://huggingface.co/qwant |
| Bureautique | Microsoft Office, Libre/Open Office, LaTeX |
| Scripts and programming | C++, Python, Perl, Awk, Bash, SQL |
| Computer Science knowledge | Natural Language Processing, Information Retrieval, Artificial Intelligence, Machine Learning, Deep Learning, classification, Graphs, Databases |

| Tools | | |
|---|---|---|
| | Machine Learning: | Keras–TensorFlow, PyTorch |
| | Speech processing: | Kaldi |
| | Machine Translation: | Marian-NMT, OpenNMT, Moses, GIZA++, MERT |
| | Language models: | SLM, LLM, SRILM, KenLM |
| | Finite State Machines: | OpenFST, AT&T's FSM & GRM library |
| | Information Retrieval: | Vespa, Lucene |

## Languages

| | |
|---|---|
| French | Native |
| English | Scientific / working |

## Industrial Highlights

- In charge of innovation transfer projects going from Technology Readiness Level (TRL) 1 to 9.
- In charge of building and managing an NLP team composed of research scientists and engineers (4 to 5 people).
- Successfully designed and developed a generic in-production framework for AI approaches based on C++ REST API (see my github).
- Successfully lead custom in-production projects for internal demands.
- Act as Deputy Chief Science Officer for 4 years, answering directly to the CTO.

## Research background

Since my thesis [Servan, 2008], understanding for human-machine interaction systems has been the common thread running through my work, with one question: how can we transpose the meaning of an utterance into a representation that is usable and meaningful for the machine?

This notion of understanding and meaning is applied in two main domains:

○ utterance analysis for a dialogue system, e.g.: entity detection (slot-filling) & intention detection.

○ extracting information from queries for a search engine, e.g.: query understanding & intention detection.

I also explored these approaches from a multilingual perspective.

Finally, my experience in the industry has made me aware of the usability of these approaches, their scaling up, and their computational, energy, and ecological costs.

Three application domains are concerned by my work on machine (and deep) learning applied to Natural Language Processing (NLP): Spoken Language Understanding, Machine Translation, and Information Retrieval.

### End-to-end approaches

My studies have focused on *end-to-end* approaches, firstly in speech understanding using finite-state automata [Servan et al., 2006]. This approach preserves all hypotheses from signals to concepts. The choice of the best hypothesis is no longer linked to the sequentiality of approaches but takes into account all possibilities from the speech signal.

In addition, my work in neural speech translation was the first to explore the possibility of end-to-end speech translation, from speech signal to text [Bérard et al., 2016].

### Domain and context adaptation for Machine Translation

I also explored multicriteria optimization [Servan and Schwenk, 2011] and domain adaptation, applied to Machine Translation. In domain adaptation, I proposed data selection approaches using language models [Lambert et al., 2011, Cettolo et al., 2013, Cettolo et al., 2014], then these approaches were put in competition with other similar approaches [Wuebker et al., 2014]. These approaches have shown the impact of selections of styles and domain-adapted lexicons, which result in improved translation results.

I proposed and applied a continuation of this work to finer contexts, which led us to use a document representation for MT in order to contextualize sentence translation [Macé and Servan, 2019]. In the latter case, the approach showed improvements in translation scores and, after analysis of the results, at lexical, syntactic, and even semantic levels.

### Evaluation methods

Having participated early on in the organization and participation of an evaluation campaign [Bonneau-Maynard et al., 2006], it was only natural that I should take part in various evaluation campaigns for machine translation [Schwenk et al., 2011, Lambert et al., 2011, Servan and Petitrenaud, 2012].

This expertise enabled me to study machine metrics in translation and to pioneer using neural representations for evaluating and augmenting machine metrics for TER and METEOR [Le et al., 2016, Servan et al., 2016].

Recently, at Qwant and within the framework of the KODICARE project (founded by the French National Research Agency – ANR), we investigated how to evaluate web search engines using a dynamic collection. To put our research to the community, we organized a CLEF evaluation lab [Alkhalifa et al., 2023].

### Usability of large-scale eco-responsible LLM

Scaling up and industrializing approaches based on deep learning was one of the major challenges of my work at Qwant. In particular, for a search engine, the notion of very large scale concerns the development of query classification approaches that can reach more than 2 000 queries per second (and nearly 5 million per day) [Maudet et al., 2019, Maudet and Servan, 2020].

This leads us to study the energy and ecological impact of our models. With the lack of a compact model in French, I've proposed creating a compact BERT model that's more ecological and responsible than its competitors but as efficient [Cattan et al., 2021]. I've recently extended this monolingual model to a multilingual model [Servan et al., 2024]. It should be noted that the energy cost of a model is also linked to its financial cost, which is an increasingly important aspect nowadays.

Following this idea, my current work leads me to investigate the prompting [Lepagnol et al., 2024] and the adapter approaches [Boucharenc et al., 2024].

## Scientific Highlights

- Release of a pre-trained transformer compact model for French called FrALBERT available on HuggingFace [Cattan et al., 2021].
- Part of the executive head of the French Natural Language Processing Association (Association pour le Traitement Automatique des Langues – ATALA) since 2020
- Part of the executive of the French Information Retrieval Association (Association Francophone de Recherche d'Information (RI) et Applications – ARIA) depuis 2019
- First end-to-end approach of spoken language translation [Bérard et al., 2016]
- Pioneering end-to-end approach of spoken language understanding [Servan and Béchet, 2008]
- Best paper award at the French Natural Language Processing Conference, TALN in 2006 [Servan and Béchet, 2006]

## Founded projects

At QWANT, as scientific manager, I took part in the calls for successfully founded projects, which are presented in the following table :

| Years | Founding | Projet name | Coordinator | Budget[†] | Project role |
|-------|----------|-------------|-------------|--------|--------------|
| 2022–2024 | France Relance | MUSCQA (24 month) | *Christophe Servan* | 212K€ (212K€) | **Projet coordinator**, and local coordinator (QWANT). |
| 2019–2024 | ANR PRCE | TextToKids (42 month) | *Delphie Baptistelli (Univ. Nanterre)* | 649K€ (335K€) | Local coordinator (QWANT). |
| 2019–2024 | ANR PRCI | KodiCare (36 month) | *Philippe Mulhem (LIG)* | 1M€ (369K€) | Local coordinator (QWANT). |
| 2019–2021 | PIA–BPI | Moteur de recherche pour l'Orientation du Secondaire au Supérieur – *MOSS* (24 month) | *Mélanie Coquelin (ONISEP)* | 1M€ (631K€) | Local coordinator (QWANT). |
| 2018–2021 | H2020 | AI4EU (36 month) | Patrick Gattelier (Thalès) | 20,5M€ (388K€) | Local coordinator (QWANT). |
| 2018–2021 | H2020 | SocialTruth (36 month) | George Koutalieris (ICCS) | 3,2M€ (458K€) | Local coordinator (QWANT). |
| 2018–2021 | PIA–BPI | ANSWER (36 month) | Christophe Servan | 2M€ (1M€) | **Projet coordinator**, and local coordinator (QWANT). |

[†]: Full project budget, Qwant's budget is specified between parenthesis.

## Scientific vulgarization

- Tutorial on Deep Learning for French at TALN conference, *TutoriAL : Apprentissage Profond pour le TAL français pour les débutants (TutoriAL: Deep Learning for French NLP tasks for beginners)* – 2023
- Keynote at Paris–Sorbonne, *TAL et Moteur de recherche, avancées et perspectives (NLP and search engines, advances and outlooks)* – 2022
- Keynote at Paris–Nanterre, *Débouchés du TAL dans l'industrie (NLP career opportinties in industry)* – 2020–2022
- Keynote at Telecom–Paris, *L'innovation dans l'industrie du TAL (NLP Innovation in industry)* – 2019–2021
- Keynote at l'ETNA, QWANT : Évolutions d'un moteur de recherche (QWANT: Search Engine Evolution) – 2020
- Keynote at GFII Forum 2017, *Elevage et évolution d'IA en traduction automatique (AI breeding and evolution in machine translation)* – 2017
- Visiting days in Laboratoire Informatique d'Avignon – 2009
- Presents scientific work in colleges – 2009
- Fête de la Science, present the Human–Machine Dialogue team – 2008

## Publications details

| | |
|---|---|
| Peer-reviewed international book chapters | 1 |
| Peer-reviewed international journals | 2 |
| Peer-reviewed international conferences | 24 |
| Peer-reviewed national (French) conferences | 10 |
| Peer-reviewed international workshops | 6 |
| workshops (Evaluations) | 5 |
| Patents | 2 |
| White papers, demonstrations | 5 |

## Publication list

### – PhD Thesis

[Servan, 2008]   Servan, C. (2008). *Apprentissage automatique et compréhension dans le cadre d'un dialogue homme-machine téléphonique à initiative mixte*. Manuscrit de thèse, Université d'Avignon et des Pays de Vaucluse.

### – Peer-reviewed international book chapters

[Semmar et al., 2018]   Semmar, N., Servan, C., Laib, M., Bouamor, D., & Marchand, M. (2018). *Extracting and Aligning Multiword Expressions from Parallel Corpora*, chapter 9, (pp. 239–268). Representation and parsing of multiword expressions: Current trends. Berlin: Language Science Press.

### – Peer-reviewed international journals

[Cettolo et al., 2014]   Cettolo, M., Bertoldi, N., Federico, M., Schwenk, H., Barrault, L., & Servan, C. (2014). Translation project adaptation for MT-enhanced computer assisted translation. *Machine Translation*, *28*, 127.

[Servan and Schwenk, 2011]   Servan, C. & Schwenk, H. (2011). Optimising Multiple Metrics with MERT. *The Prague Bulletin of Mathematical Linguistics*, (96), 109.

### – Peer-reviewed international conferences

[Lepagnol et al., 2024]   Lepagnol, P., Gérald, T., Ghannay, S., Servan, C., & Rosset, S. (2024). Small Language Models are Good Too: An Empirical Study of Zero-Shot Classification. In ***Submitted to*** *The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*.

[Alavoine et al., 2024]   Alavoine, N., Laperierre, G., Servan, C., Ghannay, S., & Rosset, S. (2024). New Semantic Task for the French Spoken Language Understanding MEDIA Benchmark. In ***Submitted to*** *The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*.

[Servan et al., 2024]   Servan, C., Ghannay, S., & Rosset, S. (2024). mALBERT: Is a Compact Multilingual BERT Model Still Worth It? In ***Submitted to*** *The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*.

[Boucharenc et al., 2024]   Boucharenc, I., Gérald, T., Ghannay, S., Servan, C., & Rosset, S. (2024). Is it necessary to go through all pre-trained Model Layers for Inference? In ***Submitted to*** *The 46th European Conference on Information Retrieval (ECIR)*.

[Alkhalifa et al., 2023]   Alkhalifa, R., et al. (2023). LongEval: Longitudinal Evaluation of Model Performance at CLEF 2023. In *proceedings of the 45th European Conference on Information Retrieval 2023 (ECIR2023)* Dublin, Ireland.

[Cattan et al., 2022b]   Cattan, O., Servan, C., & Rosset, S. (2022b). On the Usability of Transformers-based models for a French Question-Answering task. In *Joint Conference of the Information Retrieval Communities in Europe (CIRCLE) 2022* Samatan, France https://hal.archives-ouvertes.fr/hal-03701740.

[Cattan et al., 2022a]   Cattan, O., Ghannay, S., Servan, C., & Rosset, S. (2022a). Benchmarking Transformers-based models on French Spoken Language Understanding tasks. In *INTERSPEECH 2022* Incheon, South Korea https://hal.archives-ouvertes.fr/hal-03715340.

[Cattan et al., 2021]   Cattan, O., Servan, C., & Rosset, S. (2021). On the Usability of Transformers-based models for a French Question-Answering task. In *Recent Advances in Natural Language Processing, RANLP 2021* Online.

[Ghannay et al., 2020]   Ghannay, S., Servan, C., & Rosset, S. (2020). Neural Networks approaches focused on French Spoken Language Understanding: application to the MEDIA Evaluation Task. In *Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)* (pp. 2722–2727). Barcelona, Spain (Online): International Committee on Computational Linguistics https://www.aclweb.org/anthology/2020.coling-main.245.

[Deng et al., 2017]   Deng, Y., et al. (2017). SYSTRAN Purely Neural MT Engines for WMT2017. In *Proceedings of the Second Conference on Machine Translation (WMT2017)* Copenhagen, Denmark.

[Servan et al., 2016]   Servan, C., Bérard, A., Elloumi, Z., Blanchon, H., & Besacier, L. (2016). Word2Vec vs DBnary: Augmenting METEOR using Vector Representations or Lexical Resources? In *COLING 2016* Osaka, Japan.

[Bérard et al., 2016]   Bérard, A., Servan, C., Pietquin, O., & Besacier, L. (2016). MultiVec: a multilingual and multilevel representation learning toolkit for NLP. In *Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC 2016)*.

[Le et al., 2016]   Le, N.-T., Servan, C., Lecouteux, B., & Besacier, L. (2016). Better Evaluation of ASR in Speech Translation Context Using Word Embeddings. In *Interspeech 2016* San-Francisco, CA, USA.

[Wuebker et al., 2014]   Wuebker, J., et al. (2014). Comparison of Data Selection Techniques for the Translation of Video Lectures. In *Proceedings of the eleventh biennial conference of the Association for Machine Translation in the Americas (AMTA-2014)* Vancouver, Canada: AMTA.

[Cettolo et al., 2013]   Cettolo, M., Servan, C., Bertoldi, N., Federico, M., Barrault, L., & Schwenk, H. (2013). Issues in Incremental Adaptation of Statistical MT from Human Post-edits. In *MT Summit XIV Workshop on Post-editing Technology and Practice* Nice, France.

[Servan and Petitrenaud, 2012]   Servan, C. & Petitrenaud, S. (2012). Calculation of phrase probabilities for Statistical Machine Translation by using belief functions. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)* Mumbai, India.

[Semmar et al., 2011]   Semmar, N., Servan, C., Bouamor, D., & Joua, A. (2011). Using Cross-Language Information Retrieval for Machine Translation. In *Proceedings of the 5th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics* Poznań, Pologne.

[Servan et al., 2010]   Servan, C., Camelin, N., Raymond, C., Béchet, F., & De Mori, R. (2010). On the Use of Machine Translation for Spoken Language Understanding Portability. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 5330 – 5333). Dallas, Texas, United States: IEEE.

[Semmar et al., 2010]   Semmar, N., Servan, C., De Chalendar, G., Le Ny, B., & Bouzaglou, J.-J. (2010). A Hybrid Word Alignment Approach to Improve Translation Lexicons with Compound Words and Idiomatic Expressions. In *Proceedings of the 32nd Translating and the Computer Conference - ASLIB* Londres, Royaume Uni.

[Duvert et al., 2008]   Duvert, F., Meurs, M.-J., Servan, C., Béchet, F., Lefèvre, F., & De Mori, R. (2008). Semantic composition process in a speech understanding system. In *Proceedings of the 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* Las Vegas, United States.

[Barrault et al., 2008]   Barrault, L., Servan, C., Matrouf, D., Linarès, G., & De Mori, R. (2008). Frame-Based Acoustic Feature Integration for Speech Understanding.   In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008.*

[Servan and Béchet, 2008]   Servan, C. & Béchet, F. (2008). Fast call-classification system development without in-domain training data. In *Proceedings of the International Conference on Speech and Language Processing (ICSLP) Interspeech 2008* Brisbane, Australia.

[Bonneau-Maynard et al., 2006]   Bonneau-Maynard, H., et al. (2006). Results of the French Evalda-Media evaluation campaign for literal understanding.   In *Proceedings of the fifth international conference on Language Resources and Evaluation (LREC 2006)* Genes, Italy.

[Servan et al., 2006]   Servan, C., Raymond, C., Béchet, F., & Nocéra, P. (2006). Conceptual decoding from word lattices: application to the spoken dialogue corpus MEDIA. In *Proceedings of the Ninth International Conference on Spoken Language Processing (Interspeech 2006 - ICSLP)* Pittsburgh, United States.

### – Peer-reviewed national (French) conferences

[Cattan et al., 2022]   Cattan, O., Ghannay, S., Servan, C., & Rosset, S. (2022).   Étude comparative de modèles Transformers en compréhension de la parole en français.   In *34e Journées d'Etudes sur la Parole (JEP2022)* Île de Noirmoutier, France https://hal.archives-ouvertes.fr/hal-03701654.

[Maudet and Servan, 2020]   Maudet, E. & Servan, C. (2020). Conception d'un système de détection d'intention pour un moteur de recherche sur Internet. In *La conférence conjointe JEP-TALN-RECITAL 2020* NANCY, France.

[Maudet et al., 2019]   Maudet, E., Cattan, O., De Seyssel, M., & Servan, C. (2019). Qwant Research @DEFT 2019: Document matching and information retrieval using clinical cases. In *Atelier Défi Fouilles de Texte 2019* TOULOUSE, France https://hal.archives-ouvertes.fr/hal-02172582.

[Servan et al., 2016]   Servan, C., Elloumi, Z., Blanchon, H., & Besacier, L. (2016). Word2Vec vs DBnary ou comment (ré)concilier représentations distribuées et réseaux lexico-sémantiques ? Le cas de l'évaluation en traduction automatique.  In *La conférence conjointe JEP-TALN-RECITAL 2016.*

[Servan and Dymetman, 2015]   Servan, C. & Dymetman, M. (2015). Adaptation par enrichissement terminologique en traduction automatique statistique fondée sur la génération et le filtrage de bi-segments virtuels. In *La 22ème Conférence sur le Traitement Automatique des Langues Naturelles* Caen, France: ATALA.

[Servan and Petitrenaud, 2012]   Servan, C. & Petitrenaud, S. (2012). Utilisation des fonctions de croyance pour l'estimation de paramètres en traduction automatique. In *La conférence conjointe JEP-TALN-RECITAL 2012* Grenoble, France.

[Duvert et al., 2008]   Duvert, F., Meurs, M.-J., Servan, C., Béchet, F., Lefèvre, F., & De Mori, R. (2008). Composition sémantique pour la compréhension de la parole dans un cadre de dialogue.  In *Les 27e Journées d'Etudes sur la Parole (JEP)* Avignon, France.

[Servan, 2006]   Servan, C. (2006). Utilisation des transducteurs dans le décodage conceptuel : application au corpus MEDIA. In *MajecSTIC* Lorient, France.

[Servan and Béchet, 2006]   Servan, C. & Béchet, F. (2006). Décodage conceptuel et apprentissage automatique : application au corpus de dialogue Homme-Machine MEDIA.  In *La 13ème édition de la conférence sur le Traitement Automatique des Langues Naturelles (TALN 2006)* Louvain, Belgium.  Best paper Award.

[Servan et al., 2006]   Servan, C., Raymond, C., Béchet, F., & Nocéra, P. (2006). Décodage conceptuel à partir de graphes de mots sur le corpus de dialogue Homme-Machine MEDIA. In *Les XXVIes Journées d'Étude sur la Parole (JEP 2006)* Dinard, France.

### – Peer-reviewed international workshops

[Maudet et al., 2019]   Maudet, E., Cattan, O., De Seyssel, M., & Servan, C. (2019). Qwant Research @DEFT 2019: Document matching and information retrieval using clinical cases. In *Atelier Défi Fouilles de Texte 2019* TOULOUSE, France https://hal.archives-ouvertes.fr/hal-02172582.

[Deng et al., 2017]    Deng, Y., et al. (2017). SYSTRAN Purely Neural MT Engines for WMT2017. In *Proceedings of the Second Conference on Machine Translation (WMT2017)* Copenhagen, Denmark.

[Servan et al., 2012]    Servan, C., Lambert, P., Rousseau, A., Schwenk, H., & Barrault, L. (2012). LIUM's SMT Machine Translation Systems for WMT 2012. In *Proceedings of the Seventh Workshop on Statistical Machine Translation (WMT12)* Montreal, Canada.

[Schwenk et al., 2011]    Schwenk, H., Lambert, P., Barrault, L., Servan, C., Afli, H., Abdul-Rauf, S., & Shah, K. (2011). LIUM's SMT Machine Translation Systems for WMT 2011. In *Proceedings of the Sixth workshop on Statistical Machine Translation* Edinburgh, United Kingdom.

[Servan and Semmar, 2010]    Servan, C. & Semmar, N. (2010). A Hybrid Approach for Machine Translation Based on Cross-language Information Retrieval. In *Proceedings of the International Workshop on Spoken Language Translation (IWSLT 2010)* Paris, France.

### – Workshops (Evaluations)

[Cattan et al., 2021]    Cattan, O., Servan, C., & Rosset, S. (2021). On the cross-lingual transferability of multilingual prototypical models across NLU tasks. In *Meta Learning and Its Applications to Natural Language Processing, ACL2021 Workshop* Online.

[Macé and Servan, 2019]    Macé, V. & Servan, C. (2019). Using Whole Document Context in Neural Machine Translation. In *Proceedings of the 16th International Workshop on Spoken Language Translation (IWSLT'19)*.

[Bérard et al., 2016]    Bérard, A., Pietquin, O., Besacier, L., & Servan, C. (2016). Listen and Translate: A Proof of Concept for End-to-End Speech-to-Text Translation. In *NIPS Workshop on end-to-end learning for speech and audio processing* Barcelona, Spain.

[Servan et al., 2015]    Servan, C., Le, N.-T., Luong, N. Q., Lecouteux, B., & Besacier, L. (2015). An Open Source Toolkit for Word-level Confidence Estimation in Machine Translation. In *Proceedings of the 12th International Workshop on Spoken Language Translation (IWSLT'15)* Da Nang, Vietnam.

[Cettolo et al., 2013]    Cettolo, M., Servan, C., Bertoldi, N., Federico, M., Barrault, L., & Schwenk, H. (2013). Issues in Incremental Adaptation of Statistical MT from Human Post-edits. In *MT Summit XIV Workshop on Post-editing Technology and Practice* Nice, France.

[Lambert et al., 2011]    Lambert, P., Schwenk, H., Servan, C., & Abdul-Rauf, S. (2011). Investigations on Translation Model Adaptation Using Monolingual Data. In *Proceedings of the Sixth Workshop on Statistical Machine Translation* (pp. 284–293). Edinburgh, United Kingdom.

### – Patents

[Servan and Dymetman, 2016]    Servan, C. & Dymetman, M. (2016). Terminological adaptation of statistical machine translation through automatic generation of phrasal contexts for bilingual terms.

[Roux and Servan, 2016]    Roux, C. & Servan, C. (2016). Sentence generation using linguistic information.

### – White papers, demonstrations, etc.

[Servan, 2020]    Servan, C. (2020). Qwant at COLING 2020: Neural Networks approaches focused on French Spoken Language Understanding: application to the MEDIA Evaluation Task. *Qwant's BetterWeb* https://betterweb.qwant.com/en/qwant-at-coling-2020-neural-networks-approaches-focused-on-french-spoken-lar

[Portaz et al., 2019]    Portaz, M., RANDRIANARIVO, H., Nivaggioli, A., Maudet, E., Servan, C., & Peyronnet, S. (2019). *Image search using multilingual texts: a cross-modal learning approach between image and text*. Research report, Qwant Research https://hal.archives-ouvertes.fr/hal-02077556.

[Crego et al., 2016]    Crego, J. M., et al. (2016). *SYSTRAN's Pure Neural Machine Translation Systems*. Technical report, SYSTRAN http://arxiv.org/abs/1610.05540.

[Senellart et al., 2016]    Senellart, J., Servan, C., & Bou, G. (2016). *Experience Pure Neural MT*. White Paper, SYSTRAN. Vulgarisation scientifique à destination des clients de SYSTRAN.

[Servan et al., 2016]    Servan, C., Crego, J., & Senellart, J. (2016). *Domain specialization: a post-training domain adaptation for Neural Machine Translation*. Technical report, SYSTRAN.