

Yoann Dupont

*Docteur en Sciences du Langage,
informaticien de formation,
spécialisé en TAL*

[...]

[...]

France

+33 0 00 00 00 00

✉ yoa.dupont@gmail.com

📄 yoanndupont.github.io

🌐 [yoann-dupont-385b0462](https://www.linkedin.com/in/yoann-dupont-385b0462)

👤 YoannDupont

permis B



Expérience professionnelle

- Nov. 2013 **Ingénieur informatique, Expert System France.**
(4 ans) Implémentation de méthodes pour la reconnaissance des entités nommées structurées par apprentissage automatique à l'aide de CRF. Définition des besoins pour l'intégration d'un module d'apprentissage actif. Participation au projet IMM (IRT SystemX) pour intégrer un module d'extraction de relations entre entités nommées : génération de traits pour le module MultiR, basés sur parsing en dépendances syntaxiques (MaltParser entraîné sur le corpus anglais de Universal Dependencies).
- Oct. 2012 **Ingénieur d'étude, Université Pierre et Marie Curie.**
(3 mois) Intégration d'un module de reconnaissance des entités nommées par apprentissage automatique (CRF). Module intégré dans SEM, logiciel libre open source, de ma création.
- Avr. 2012 **chercheur stagiaire (master), Laboratoire Lattice.**
(6 mois) Combinaisons de méthodes symboliques (généralisation d'automates) et d'apprentissage statistique (CRF) appliquées à la tâche de chunking syntaxique.
- Avr. 2010 **chercheur stagiaire (licence), Laboratoire d'Informatique Fondamentale d'Orléans.**
(3 mois) Intégration de connaissances dans un CRF pour l'annotation morphosyntaxique.

Expérience d'enseignement

- Jan. 2016 **chargé de travaux dirigés, Université Sorbonne Nouvelle Paris 3.**
(4 mois) Licence 2 Sciences du Langage, module « Informatique et Industries de la Langue », encadré par Serge Fleury, Kim Gerdes et Isabelle Tellier. 1 groupe de 20 étudiants. Durée du TD : 18h.
- Jan. 2015 **chargé de travaux dirigés, Université Sorbonne Nouvelle Paris 3.**
(4 mois) Licence 2 Sciences du Langage, module « Informatique et Industries de la Langue », encadré par Serge Fleury, Kim Gerdes et Isabelle Tellier. 1 groupe de 20 étudiants. Durée du TD : 18h.
- Sept. 2013 **chargé de travaux dirigés, Université Sorbonne Nouvelle Paris 3.**
(4 mois) Licence 2 Sciences du Langage, module « Informatique et Industries de la Langue », encadré par Serge Fleury, Kim Gerdes et Isabelle Tellier. 3 groupes de 10 à 20 étudiants. Durée du TD : 9h.
- Jan. 2012 **Tuteur, Université d'Orléans.**
(4 mois) Tuteur pour les étudiants de première année de licence informatique en difficulté.

Formation

- 2013 – 2017 **Thèse CIFRE, Université Sorbonne-Nouvelle Paris 3 et Expert System France,**
soutenue le 23 / 11 / 2017.
« La structuration dans les entités nommées ». Extraction de traits pertinents pour améliorer la reconnaissance des entités nommées, développement de techniques pour récupérer la structure interne des entités nommées, comparaison de méthodes par apprentissage (CRF et réseaux de neurones). Développement d'un système état-de-l'art sur le French Treebank en entités nommées.
- 2010 – 2012 **Master d'informatique, Université d'Orléans.**
Spécialité IRAD (Informatique, Répartition et Aide à la Décision)
- 2006 – 2010 **Licence d'informatique, Université d'Orléans.**
Spécialité STIC (Sciences et Technologies de l'Information et de la Communication)

Compétences

| | |
|--------------|---|
| Langages | Python (bibliothèque TAL), Java (module de CRF), Ada (patrons de conceptions), C/C++, Lua (module Torch), Latex, bash, PowerShell, OCaML, Prolog |
| Librairies | Torch, Theano (notions), Wapiti, MultiR, MaltParser |
| Systèmes | Windows (7, 10) , Linux (Ubuntu, Mint, Debian) |
| Outils | GIT, maven, JIRA, Eclipse, Microsoft Office, LibreOffice |
| Méthodologie | Patrons de conceptions, programmation générique, Scrum |

Langues

| | Comprendre | | Parler | | Écrire | |
|----------|------------|------|-----------|------------|--------|--------------------------|
| | Écouter | Lire | Converser | S'exprimer | | |
| Français | C2 | C2 | C2 | C2 | C2 | <i>langue maternelle</i> |
| Anglais | B2 | C1 | B2 | B2 | C1 | |

Projets libres open source

- SEM Segmenteur Étiqueteur Markovien. Bibliothèque python pour l'annotation. Bibliothèque créée durant le stage de licence et enrichie régulièrement depuis. Permet l'annotation PoS, le chunking et en entités nommées. Export en plusieurs formats pour interopérabilité. [\[github\]](#)
- GAAL Generic Ada Automata Library. Bibliothèque d'automates codée en Ada. Implémente notamment l'algorithme k-RI. Bibliothèque créée durant le stage de master. [\[github\]](#)

Références

- Isabelle Tellier Professeure des Universités au Laboratoire Lattice. Contact : isabelle.tellier@sorbonne-nouvelle.fr
- Marco Dinarelli Chargé de Recherches au Laboratoire Lattice. Contact : marco.dinarelli@ens.fr
- Christian Lautier Directeur Technique à Expert System France. Contact : clautier@expertsystem.com

Publications

Articles de revue avec comité de lecture

DINARELLI, Marco et DUPONT, Yoann.

Modélisation de dépendances entre étiquettes dans les réseaux neuronaux récurrents.

In: Revue TAL. 2017, vol. 58, no 1. (*accepté*)

Conférences internationales

DUPONT, Yoann, DINARELLI, Marco, TELLIER, Isabelle and LAUTIER, Christian.

Structured Named Entity Recognition by Cascading CRFs.

In: CICling. 2017.

[article](#)

DUPONT, Yoann, DINARELLI, Marco and TELLIER, Isabelle.

Label-Dependencies Aware Recurrent Neural Networks (Prix du meilleur programme CICling 2017).

In: CICling. 2017.

[article](#)

TELLIER, Isabelle, MAKHLOUF, Zineb and DUPONT, Yoann.

Sequential Patterns of POS Labels Help to Characterize Language Acquisition.

In: DMNLP @ PKDD/ECML. 2014. p. 129-142.

[article](#)

TELLIER, Isabelle, DUPONT, Yoann, ESHKOL-TARAVELLA, Iris and WANG, Ilaine.

Adapt a Text-Oriented Chunker for Oral Data: How Much Manual Effort Is Necessary?

In: IDEAL. 2013. p. 226-233.

[article](#)

TELLIER, Isabelle and DUPONT, Yoann.

How Symbolic Learning Can Help Statistical Learning (and vice versa).

In: RANLP 2013. p. 649-658.

[article](#)

Conférences nationales

DUPONT, Yoann.

Exploration de traits pour la reconnaissance d'entités nommées du Français par apprentissage automatique (Prix du meilleur article RECITAL 2017**).**

In: TALN-RECITAL. 2017.

[article](#)

DUPONT, Yoann, DINARELLI, Marco et TELLIER, Isabelle.

Réseaux neuronaux profonds pour l'étiquetage de séquences.

In: TALN-RECITAL. 2017.

[article](#)

DUPONT, Yoann, TELLIER, Isabelle, LAUTIER, Christian, et DINARELLI, Marco.

Extraction automatique d'affixes pour la reconnaissance d'entités nommées chimiques.

In: EGC. 2016.

[article](#)

MAKHLOUF, Zineb, DUPONT, Yoann, et TELLIER, Isabelle.

Caractériser l'acquisition d'une langue avec des patrons d'étiquettes morpho-syntaxiques.

In: JADT. 2014.

[article](#)

TELLIER, Isabelle, DUPONT, Yoann, ESHKOL-TARAVELLA, Iris et WANG, Ilaine.

Peut-on bien chunker avec de mauvaises étiquettes POS ?

In: TALN. 2014. p. 125-136.

[article](#)

TELLIER, Isabelle et DUPONT, Yoann.

Apprentissage symbolique et statistique pour le chunking : comparaison et combinaisons.

In: TALN-RECITAL 2013.

[article](#)

CONSTANT, Matthieu, TELLIER, Isabelle, DUCHIER, Denys, DUPONT, Yoann, Anthony Sigogne, Sylvie Billot
Intégrer des connaissances linguistiques dans un CRF: application à l'apprentissage d'un segmenteur-étiqueteur du français.

In: TALN 2011. p. 321.

[article](#)

Démonstrations

DUPONT, Yoann et PLANCQ, Clément.

Un étiqueteur en ligne du Français.

session démonstration de TALN-RECITAL, 2017, p. 15.

[article](#)

DUPONT, Yoann et TELLIER, Isabelle.

Un reconnaiseur d'entités nommées du Français.

. 2014. p. 40-41.

[article](#)

TELLIER, Isabelle, DUPONT, Yoann et COURMET, Arnaud.

Un segmenteur-étiqueteur et un chunker pour le français.

In: JEP-TALN-RECITAL 2012.

[article](#)

Centres d'intérêt

- Sport
- Cuisine
- Pâtisserie
- Musique
- Vidéo
- Langues