



**HAL**  
open science

## La préparation des données pour une typologie et une ontologie de l'expression du déplacement

Karl Seifen, Carmelo Alessandro Basile, Coline Caillol, Cameron Morin,  
Moisés Velásquez, Chenyang Zhao

► **To cite this version:**

Karl Seifen, Carmelo Alessandro Basile, Coline Caillol, Cameron Morin, Moisés Velásquez, et al.. La préparation des données pour une typologie et une ontologie de l'expression du déplacement. Étudier le Langage à l'Ère Numérique - Actes des 25èmes Rencontres des Jeunes Chercheurs (RJC) en Sciences du Langage, 2023. hal-04894461

**HAL Id: hal-04894461**

**<https://hal.science/hal-04894461v1>**

Submitted on 23 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

La préparation des données pour une typologie et une ontologie de l'expression du  
déplacement

*Karl Seifen*

Dynamique du Langage, Université Lumière – Lyon 2

[karl.seifen@univ-lyon2.fr](mailto:karl.seifen@univ-lyon2.fr)

**RESUME**

Cet article a pour but de présenter la méthodologie utilisée dans le cadre d'une étude en cours sur l'expression du déplacement. Nous montrons comment des données de langues diverses typologiquement et géographiquement peuvent être formatées et codées pour créer une ontologie (un réseau de concepts). Le codage des données se fait à deux niveaux : le premier est morphosyntaxique (propriétés syntaxiques et morphologiques des formes linguistiques) et le second sémantique (catégories conceptuelles relevant du domaine spatial). Nous illustrons notre méthodologie par une recherche sur le type d'unités encodant le concept sémantique de Trajectoire dans nos données et montrons que différentes langues encodent la Trajectoire dans la Tête de la phrase (prédicat), d'autres dans un Relateur (cas ou adposition), ou encore dans de multiples *loci* : Tête et Relateur, Tête et Satellite (particule ou affixe verbal), ou Relateur et Satellite.

*Mots-clés* : *Typologie de l'espace – Trajectoire – Ontologie – Déplacement*

**ABSTRACT**

This paper aims to show the methodology employed in an on-going study of motion expression. We show how data from typologically and geographically diverse languages can be formatted and annotated to create an ontology (a network of concepts). Data are annotated at two different levels: (i) morpho-syntactic (syntactic and morphological properties of the linguistic units) and (ii) semantic (conceptual categories of the spatial domain). We illustrate our methodology with a query on the types of units encoding the semantic concept of Path in our data and show that language families encode Path in the Head of the sentence (predicate), in a Relator (case or

adposition) or in multiple loci: Head and Relator, Head and Satellite (particle or verbal affix), or Relator and Satellite.

**Keywords** : *Typology of space –Path – Ontology – Motion*

Notre étude porte sur l'expression du déplacement dans le cadre du projet SpOTy (Spatial Ontology and Typology, Labex ASLAN, Université de Lyon)<sup>5</sup>. Ce projet vise à créer d'une part une typologie de l'expression de la Trajectoire dans les langues du monde, et d'autre part une ontologie des relations spatiales. Cet article a pour but de présenter notre méthodologie. Plus précisément, nous montrons comment le formatage et le codage des données nous permet de créer un réseau de concept interconnecté (*i.e.* une ontologie) que nous pouvons interroger. Dans un premier temps, nous définissons les concepts d'ontologie (section 1.) et d'évènement de déplacement (section 2.). Nous présentons ensuite la méthodologie employée pour mettre en forme et coder nos données (section 3.). Enfin, nous formulerons une typologie préliminaire de la Trajectoire à partir de notre base de données (section 4.).

## 1. ONTOLOGIE

Dans le domaine de l'intelligence artificielle et de la représentation des connaissances, une ontologie est un réseau interconnecté de concepts et de relations entre ces concepts, explicitant et organisant un domaine spécifique (Guarino *et al.* 2009: 2, Schalley *et al.* 2014: 141). Gruber (1993: 199) définit une ontologie comme la « spécification explicite d'une conceptualisation », *i.e.* une certaine vision du monde ou d'un domaine d'intérêt, dont les concepts sont explicitement donnés (Sowa 2003).

« A body of formally represented knowledge is based on a conceptualization: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them. A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to

---

<sup>5</sup> L'auteur remercie le LABEX ASLAN (ANR-10-LABX-0081) de l'Université de Lyon pour son soutien financier dans le cadre du programme français "Investissements d'Avenir" géré par l'Agence Nationale de la Recherche (ANR).

some conceptualization, explicitly or implicitly. » (Genesereth & Nilsson 1987, cités par Guarino *et al.* 2009: 3)

En linguistique, et plus particulièrement en typologie, les ontologies facilitent la gestion et l'analyse de données translinguistiques, qui peuvent être complexes, éparses, ou incomplètes (Schalley 2019 : 11). Les ontologies peuvent également servir à combler le manque d'interopérabilité entre différents projets et bases de données (Farrar & Langendoen 2003: 100).

Concrètement, les ontologies sont des réseaux où différents nœuds (concepts) sont reliés entre eux par des relations spécifiques. La figure 1 schématise le concept d'ontologie tel qu'il peut être utilisé en linguistique, avec comme exemple les clitiques et affixes pronominaux du français. On y retrouve les concepts de l'ontologie (encadrés), divisé en trois types : (i) les formes (*i.e.* signifiants saussuriens : 'je', 'tu', 'nous', '-ons', '-ez'), (ii) les sens (*i.e.* signifiés : singulier, pluriel, 1ère personne, 2ème personne), et (iii) les catégories morphosyntaxiques (clitiques et suffixes). Ces concepts sont reliés par des relations spécifiques (en italique) : les formes et leurs sens par la relation *a le sens de*, ainsi que les formes et leurs catégories par la relation *appartient à la catégorie morphosyntaxique*. Les concepts de même type peuvent être reliés par la relation hyperonymique *type de* : un suffixe est un type d'affixe, le singulier et le pluriel sont des instances de nombre grammatical, et les première et deuxième personnes sont des instances de personne grammaticale.

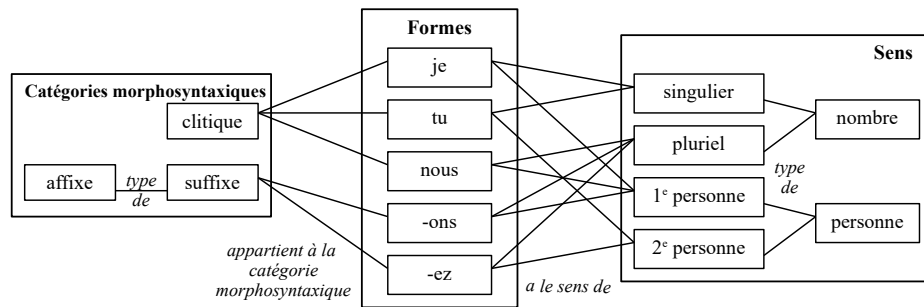


Figure 9 : Ontologie des formes pronominales du français

## 2 EXPRESSION DU DEPLACEMENT

Notre étude porte sur l'expression du déplacement et les événements dits translocatifs (aussi appelés : cislocatifs, *motion events*, *translatory situations*, *displacement*). Ce type d'évènement

exprime un changement de localisation d'une entité. Talmy (1972: 1) définit les événements translocatifs comme le déplacement ou la localisation d'un objet par rapport à un autre objet<sup>6</sup>.

Outre une définition des événements translocatifs, Talmy (1972, 1985, 2000) propose également une décomposition de ces événements en constituants sémantiques (section 2.1), ainsi qu'une typologie des langues selon leur encodage du déplacement (section 2.2).

### 2.1 Constituants des événements translocatifs

Talmy (1972, 1985) propose d'analyser ces événements translocatifs comme contenant quatre constituants principaux : (i) la Figure, *i.e.* l'entité se déplaçant, (ii) le Site (*Ground*), *i.e.* l'entité par rapport à laquelle la Figure se déplace, (iii) le Moteur (*Fact-of-Motion*), *i.e.* l'état de la Figure, pouvant avoir les valeurs dynamique (déplacement) ou statique (localisation), et (iv) la Trajectoire (*Path*), *i.e.* la relation spatiale entre la Figure et le Fond. Outre ces quatre constituants, les événements translocatifs peuvent également incorporer des co-événements comme la Manière ou la Cause du mouvement (exemple 1).

- |            |                |             |           |
|------------|----------------|-------------|-----------|
| 1. The man | ran            | into        | the room. |
| Figure     | Manière/Moteur | Trajectoire | Site      |

Toujours selon Talmy (2000: 53-56), le composant de Trajectoire est lui-même complexe et se décompose en trois sous-éléments principaux :

- 1) le Vecteur qui correspond à la position du Site sur le parcours de la Figure (point initial, point médian, étendue médiane, point final).
- 2) la Conformation qui indique la relation topologique entre la Figure et le Site (dans, sur, *etc.*)<sup>7</sup>.
- 3) la Deixis qui précise la direction du déplacement par rapport à un centre déictique (s'en approchant ou s'en éloignant). Son statut de sous-composant de la Trajectoire, plutôt que de composant constitutif des événements translocatifs, est cependant discuté (Morita 2011: §72, Matsumoto *et al.* 2017: 121, Seifen *et al.* 2021).

Outre ces trois sous-composants, Slobin et Hoiting (1994: 498) en se basant sur les travaux de Aske (1989), montrent l'importance du composant sémantique du Franchissement de frontière (ex. « entrer », « sortir », et « traverser » par opposition à « atteindre », « quitter », et « passer »).

---

<sup>6</sup> « One object moves or is located with respect to another object » (Talmy 1972:1).

<sup>7</sup> Vandeloise (1986: 22-30) montre que les prépositions spatiales du français telles que « dans » ont des caractéristiques fonctionnelles plutôt que topologiques.

## 2.2 Typologie(s) de la Trajectoire

A partir de sa description des événements translocatifs, Talmy (1985) propose une typologie des langues, selon le *locus* du composant Trajectoire, c'est-à-dire selon l'unité morphosyntaxique encodant la Trajectoire. Deux types de langues sont ainsi décrits : les langues à cadrage verbal (*verb-framed languages*) et les langues à cadrage satellitaire (*satellite-framed languages*). Les langues à cadrage verbal encodent la Trajectoire dans le verbe (langues romanes, sémitiques, polynésiennes). Ainsi dans l'exemple 2a, la Trajectoire est encodée dans le verbe *entrer*. Les langues à cadrage satellitaire (langues germaniques, slaves, sinitiques) encodent quant à elles la Trajectoire dans un satellite. Celui-ci peut être défini comme un constituant de surface, autre qu'un syntagme nominal ou prépositionnel, dépendant du verbe (Talmy 1985: 102), incluant les particules verbales, les affixes verbaux, ou les noms incorporés<sup>8</sup>. L'exemple 2b illustre ce deuxième type : la particule verbale *in* porte l'information de Trajectoire.

2. Cadrages verbal et satellitaire :

- a. L'enfant **entra**. (cadrage verbal)
- b. The child walked **in**. (cadrage satellitaire)

Cette typologie a fait l'objet de nombreuses révisions. Wälchli (2001: 301) propose trois types de langues : verbale (la Trajectoire est encodée dans un verbe), adverbiale (la Trajectoire est encodée dans une particule ou un affixe verbal) et adnominale (la Trajectoire est encodée dans un cas ou une adposition)<sup>9</sup>. Matsumoto (2003: 410) propose de remplacer les notions de verbe et de satellite par celle de tête et de non-tête (*head* et *non-head*), d'une part parce que les verbes non-finis (gérondifs, connectifs, participes) ont un comportement semblable aux satellites, et d'autre part parce que les adpositions et les cas ne sont pas des satellites.

« Talmy's typology of verb vs. satellite-framed languages suffers from the misleading use of the term "verb". What is meant by the term verb is in fact the head of a clause. Satellites can also be a verb [...]. For this reason, a better name for verb-framed languages is head-framed languages. Satellite-framed languages, on the other hand, can be termed as nonhead-framed languages. Note that satellites and nonheads are slightly

---

<sup>8</sup> Bien que la définition que donne Talmy du satellite exclue les adpositions, ces dernières sont paradoxalement incluses dans le cadrage satellitaire (voir Imbert *et al.* (2011) sur la problématique de la notion de satellite dans les travaux de Talmy).

<sup>9</sup> « According to the locus of highest differentiation of displacement one may distinguish three gross types for the encoding of displacement: (V) **verbal encoding** (i.e. by the verb stem), (AN) **adnominal encoding** (i.e. by prepositions, postpositions or case marking), and (AV) **adverbal encoding** (i.e. by verb affixes or verb particles) » (Wälchli 2001: 301).

different notions: all satellites are nonheads by definition [...] but not all nonheads are satellites. » (Matsumoto 2003: 410)

Enfin, Fortis et Vittrant (2011, 2016)<sup>10</sup> proposent une typologie des constructions, plutôt que des langues (les langues possédant généralement plusieurs types de constructions). Celle-ci est basée sur les différents *loci* possibles de la Trajectoire, à savoir :

- 1) la Tête (*i.e.* le prédicat, qu'il soit verbal ou non comme dans l'exemple 3a),
- 2) le Satellite (particule, affixe verbal, adverbial),
- 3) le Relateur (cas ou adposition),
- 4) l'Argument (généralement un nom déverbal, comme en 3b),
- 5) la construction elle-même ; la Trajectoire n'est alors pas portée par un marqueur mais par la construction elle-même, comme dans l'exemple 3c. Dans cet énoncé, le sens de « traverser » est véhiculé par la construction plutôt que par un morphème spécifique),
- 6) plusieurs *loci* possibles (combinaison de plusieurs stratégies, comme dans l'exemple 3d).

« Nous avons choisi de répartir les éléments susceptibles d'exprimer la trajectoire en quatre catégories. Ces quatre catégories sont : la tête de la phrase (notée T), le satellite de la tête [...] ou du nom, noté S, les adnominaux (A), terme qui subsume les cas et adpositions, et enfin les noms (N). Le cadrage multiple montre que ces loci d'encodage sont cumulables dans bon nombre de langues. » (Fortis & Vittrant 2011: 83)

3. a. Indonésien (Malayo-polynésien, austronésien, Indonésie) : Tête non-verbale

Məreka **kə** bioskop

3PL **to** movies

'They (are going) to the movies.' (Fortis & Vittrant 2016: 358)

- b. Français (Roman, indo-européen, France) : Argument et Relateur

L'**acheminement jusqu'à** l'aéroport se fait en navette. (Fortis & Vittrant 2016: 355)

- c. Japonais (Japonique, Japon) : Construction et Tête verbale

Taroo-ga Omekaidoo-o hashit-te it-ta.

Taro-NOM Ome\_road-ACC run-CVB go-PST

'Taro ran across the road.' (lit. Taro went the road running) (Morita 2009: 237, cité par Fortis & Vittrant 2016: 368)

---

<sup>10</sup> Dans le cadre du projet de recherche Typologie de la Trajectoire (2006-2011, Fédération Typologie et Universaux du Langage).

d. Japonais (Japonique, Japon) : Tête, Satellite et Relateur

onnanohito-ga dookutu-no naka-**kara** **de**-te **iki**-masi-ta  
woman-NOM cave-GEN inside-ABL **exit**-CVB **go**-POL-PST

‘The woman exited from the inside of the cave away from me.’ (Fortis & Vittrant 2016: 367)

### 3. DONNEES

Notre étude vise à produire une typologie de la Trajectoire dans les langues du monde. Nous avons donc compilé des données décrivant des événements translocatifs. Notre corpus actuel est constitué d'énoncés de 57 langues appartenant à 20 familles différentes. Les données sont issues de deux sources principales : (i) la littérature scientifique traitant de l'espace, et plus particulièrement de l'espace dynamique (déplacement), et (ii) les corpus personnels des membres du projet 'Typologie de la Trajectoire' (2006-2011, Fédération Typologie et Universaux du Langage), ainsi que des chercheurs appartenant aux laboratoires associés.

Les données collectées au sein du projet Typologie de la Trajectoire l'ont été à l'aide du stimulus Trajectoire (Ishibashi *et al.* 2006, Vuillermet & Kopecka 2019)<sup>11</sup>. Ce stimulus a pour but d'éliciter des descriptions de déplacement dans des langues typologiquement différentes, en demandant aux informateurs de décrire différentes scènes. L'utilisation du stimulus Trajectoire permet d'avoir des données comparables entre différentes langues. Il contient 76 clips vidéo d'une dizaine de secondes (2 entraînements, 9 distracteurs, 65 trajectoires). Les exemples 4a (mandarin), 4b (ese ejja), et 4c (polonais) décrivent tous la même situation : une femme sort d'un champ de maïs (figure 2).

---

<sup>11</sup> Disponible en ligne : <http://tulquest.huma-num.fr/en/node/132>.





Figure 10 : Clip 38 du stimulus Trajectoire (038\_Path\_F\_walk\_outof\_field\_sideRL)

4. a. Chinois mandarin (Sinitique, sino-tibétain ; Chine)

yí wèi nǚshì zǒu=chu yùmǐ dì  
 un CLF femme marcher=OUT maïs champ

‘Une femme sort du champ de maïs.’ (Jin-Ke Song, corpus personnel)

b. Ese Ejja (Tacana ; Bolovie)

epona kwaya'yoani shixeiye=xeshaja mexi  
 femme sort champ.de.maïs=PERL avec un panier

‘Une femme sort du champ de maïs avec un panier.’ (Vuillermet 2012: 574)

c. Polonais (Slave, indo-européen ; Pologne)

kobieta wy-chodzi z pola kukurydzy  
 femme OUT-marche de champ.GEN maïs.GEN

‘Une femme sort du champ de maïs en marchant.’ (Fagard & Kopecka 2021:

161)

### 3.1 Formatage

Les données collectées auprès de différentes sources dans le cadre de notre étude doivent être formatées, afin de permettre leur intégration dans une base de données (dont la création est un de nos objectifs). Ce formatage suit deux contraintes : d’une part, il doit être lisible par la base de données qui en récupère les informations, et d’autre part, il doit être suffisamment intuitif pour permettre, à terme, aux chercheurs de formater et d’intégrer eux-mêmes leurs données dans la base.

Nous avons donc opté pour un format de type tableur balisé, où les données et les métadonnées sont divisées en ligne, chacune introduite par des balises (*i.e.* mots-clés). Les balises et les informations correspondantes sont données dans le tableau 1.

Balise	Information
ID	Identifiant unique de l'énoncé
LGE	Langue de l'énoncé (suivant les conventions ISO 639-3)
TRANS	Transcription de l'énoncé
ORTHO	Enoncé en orthographe native (si applicable)
TRAD	Traduction de l'énoncé en anglais
TRADFR	Traduction de l'énoncé en français (si disponible)
TRAJ	Numéro du stimulus Trajectoire associé à l'énoncé (si applicable)
SOURCE	Référence bibliographique ou corpus personnel
TOKEN	Transcription tokénisée de l'énoncé
GLOSE	Ligne de glose interlinéaire
CLAUSE	Division de l'énoncé en propositions
MS	Codage morphosyntaxique (cf. infra)
CC	Codage des catégories conceptuelles (cf. infra)

Tableau 1: Balises et informations correspondantes

Les énoncés sont associés à un identifiant (ID), à une langue au format ISO (LGE), à une référence bibliographique ou un corpus personnel (SOURCE) et éventuellement à une scène issue du stimulus Trajectoire (TRAJ). Chaque énoncé contient également une transcription (TRANS), une traduction en anglais (TRAD), voire dans d'autres langues de travail (français : TRADFR, espagnol : TRADES), ainsi qu'une version en orthographe de la langue, si la langue est écrite (ORTHO). Les données elles-mêmes contiennent une transcription tokénisée en morphèmes (TOKEN), accompagnée pour chaque morphème d'une glose interlinéaire (GLOSE), du numéro de la proposition à laquelle ils appartiennent (CLAUSE), et d'informations morphosyntaxiques (MS) et sémantiques (CC) lorsque le morphème relève du domaine spatial. Ce formatage est illustré en figure 3, avec un exemple de thaï central (tai-kadaï, Thaïlande).

ID	tha_50						
LGE	THA						
TRANS	phǔːjǐŋ	dɤːn	kʰŭn	paj	bon	nɤːn	
ORTHO	ผู้หญิงเดินขึ้นไปบนเนิน						
TRADFR	La femme monte sur le monticule [en marchant en s'éloignant du CD].						
TRAD	The woman walks up the knoll [away from DC].						
TRAJ	71						
SOURCE	Seifen p.c.						
TOKEN	phǔːjǐŋ	dɤːn	kʰŭn	paj	bon	nɤːn	
GLOSE	woman	walk	ascend	go	top	knoll	
CLAUSE		1	1	1	1	1	1
MS	An	Hv	Hv	Hg	Rn	An	
CC	Fh	Mm	Pvu	Ddf	Cto	Gg	

Figure 11 : Exemple de formatage des données

### 3.2 Codage

Chaque morphème relevant de la spatialité reçoit deux codes. Le premier rend compte de ses propriétés morphosyntaxiques (ligne MS) et le second de ses propriétés sémantiques (ligne CC, catégorie conceptuelle). Le codage de ces informations a plusieurs avantages. Tout d'abord, il permet de normaliser la notation des éléments spatiaux issus de sources diverses. Par exemple, le sens de « déplacement à l'intérieur de » pouvant être glosé par *enter*, *go\_in*, *in*, ou *into* (sans compter les gloses non-anglaises) est rendu par « Pbi » (Trajectoire avec franchissement de frontière vers l'intérieur du Site, en anglais : *Path with boundary-crossing into a Ground*) dans notre codage, quelle que soit la glose utilisée par l'auteur d'origine.

D'autre part, le codage utilisé participe à une ontologie, où les formes linguistiques sont ainsi liées à des catégories de deux types : (i) catégorie sémantique avec une relation de type *a le sens de X* et (ii) catégorie morphosyntaxique avec une relation de type *a la propriété morphosyntaxique X*. Le codage choisi note également des relations hyperonymiques (concepts emboîtés). Ainsi, le code « Pbi » ('déplacement à l'intérieur de') est un sous-type de « Pb » (déplacement avec franchissement de frontière : *boundary*), lui-même sous-type de « P » (Trajectoire : *Path*).

Les codages morphosyntaxiques ont donc la forme « Xy(z) ». La première lettre du codage (« X ») indique la fonction syntaxique de l'unité : Tête (prédicat), Satellite (particules et affixes verbaux), Relateur (cas et adpositions), Argument, ou Satellite de l'argument (modifieur du nom). 'H' indique la Tête, 'S' le Satellite, 'R' un Relateur et 'A' un Argument et 'Z' un Satellite de l'argument. La deuxième lettre (« y ») indique la catégorie morphosyntaxique : verbe fini (v), verbe non-fini (f), nom (n), pronom (p), affixe ou clitique (x),

adposition (a), etc. Enfin, la troisième lettre (« z ») est optionnelle et marque une catégorie grammaticale pertinente pour notre travail comme la diathèse (e) ou mouvement associé (m)<sup>12,13</sup>. Les trois éléments du code morphosyntaxique « X », « y », et « z » sont indépendants. Il n'existe aucune limite combinatoire pour l'ontologie (mais en réalité certaines combinaisons semblent improbables).

Le deuxième codage (CC : catégories conceptuelles) est sémantique et de type « Abc », où « c » est un type de « Ab », lui-même type de « A ». Par exemple, la catégorie « D » (Deixis) possède les sous-types « Ds » (Deixis statique : *ici, là-bas*) et « Dd » (Deixis dynamique : déplacement par rapport au centre déictique), lui-même possédant les sous-types « Ddf » (centrifuge : s'éloignant du centre déictique *en allant*), « Ddp » (centripète : s'approchant du centre déictique *en venant*), et « Ddt » (transverse : sans s'approcher ni s'éloigner du centre déictique, *en passant*), comme montré en tableau 2. Le codage des catégories conceptuelles concerne l'ensemble des constituants sémantiques présentés en 2.1. (Figure, *Fact-of-Motion*, Site, Manière, Cause, Trajectoire et ses sous-composants : Vecteur, Conformation, Deixis).

Niveau 1	Niveau 2	Niveau 3
« D » : Deixis	« Dd » : Deixis dynamique	« Ddf » : centrifuge
		« Ddp » : centripète
		« Ddt » : transverse
	« Ds » : Deixis statique	

Tableau 2 : Codage sémantique de la Deixis

La figure 4 exemplifie le codage des propriétés morphosyntaxiques (en bleu) et sémantiques (en vert) des morphèmes de la phrase *La femme sort de la grotte*. Ici, *femme* est un Argument nominal (codage morphosyntaxique : « An ») désignant une Figure humaine (codage sémantique : « Fh »). Le verbe *sort* est la Tête de la phrase et un verbe fini (« Hv ») ; il décrit une Trajectoire avec franchissement de frontière vers l'extérieur (« Pbo »). La préposition *de* est un Relateur de type adposition (« Ra ») ; elle indique un déplacement depuis un point initial,

<sup>12</sup> Le mouvement associé (Guillaume & Koch 2021: 3) indique que l'action exprimée par le verbe est associée à un déplacement, qui peut la précéder, la suivre, ou lui être concomitant.

<sup>13</sup> Un tel marquage permet de retrouver et d'interroger les morphèmes concernés plus facilement. Ces deux catégories nous ont semblé, lors de la conception du codage, comme particulièrement intéressantes à faire ressortir ; le mouvement associé étant un sujet d'actualité, et la diathèse étant relativement peu décrite dans les études sur la spatialité.

une source (« Ps »). Enfin, *grotte* est un Argument nominal (« An ») désignant l'entité servant de référence au déplacement, ici le point initial (« Gs ») de ce déplacement.

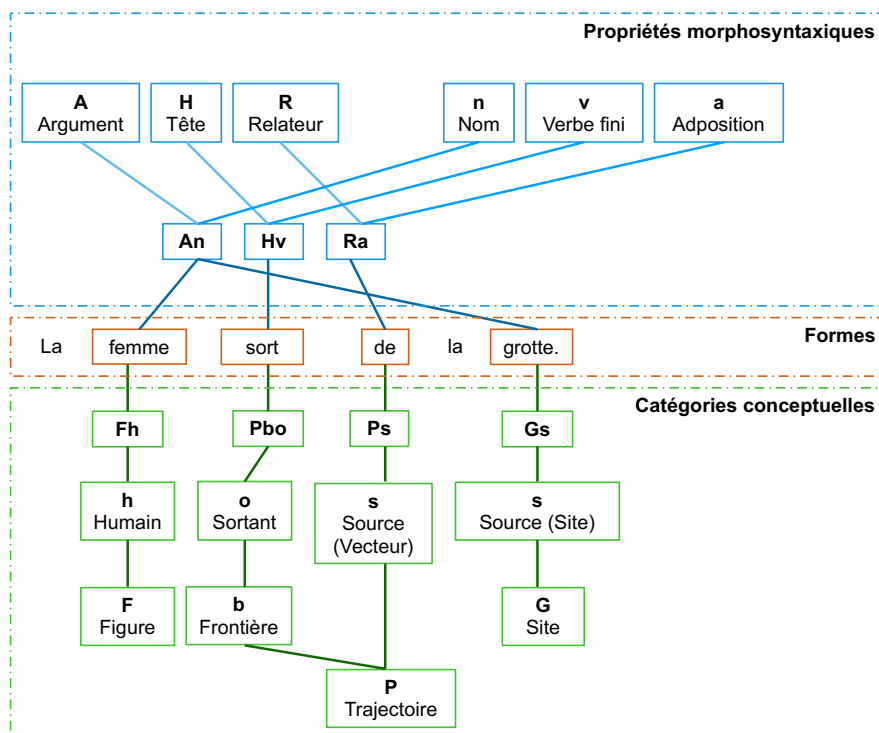


Figure 12 : Exemple de codage morphosyntaxique et sémantique

#### 4. ANALYSE PRELIMINAIRE

Le codage des données permet leur interrogation sous plusieurs angles. Il est par exemple possible de rechercher les correspondances entre catégories conceptuelles et catégories morphosyntaxiques, *i.e.* comment est encodé un concept spécifique, ou à l'inverse, qu'encode une forme particulière. Dans cette section, nous explorons l'encodage du composant sémantique de Trajectoire, dans la lignée des travaux présentés en 2.2.

Concrètement, nous recherchons dans les données les unités avec le code « P » (Trajectoire) en catégorie conceptuelle (quels que soient les sous-types), et examinons tous les codes morphosyntaxiques correspondants. Nous obtenons le résultat schématisé dans le tableau 3, où chaque ligne correspond à une unité encodant de la Trajectoire. Pour chaque unité, le tableau indique l'énoncé d'origine, la langue et le code morphosyntaxique correspondant (la forme linguistique elle-même n'est pas indiquée). Le tableau 3 est limité aux deux premières

unités de notre recherche (sur un total de 821 occurrences), qui appartiennent aux énoncés aho\_01 et aho\_02 de la langue AHO (Ahom, famille Tai-Kadaï) et codées « Hv » (tête verbale).

sentence	language	ms_code
aho_01	aho	Hv
aho_02	aho	Hv
...		

Tableau 3 : Résultats de la recherche des unités encodant la Trajectoire

Les données ainsi obtenues nous permettent d'établir le tableau 4, qui indique pour chaque sous-famille de langues la proportion de *loci* syntaxiques (têtes, relateurs, et satellites) encodant le composant sémantique de Trajectoire.

	Tête	Relateur	Satellite	
<b>Afro-Asiatique, Tchadien</b>	0,81	0,19	0,00	
<b>Austro-Asiatique, Bahnarique</b>	0,63	0,37	0,00	
<b>Austro-Asiatique, Viet-Muong</b>	1,00	0,00	0,00	
<b>Austronésien, Océanique</b>	0,27	0,60	0,13	
<b>Coréanique</b>	0,17	0,50	0,33	
<b>Hmong-Mien, Hmongique</b>	1,00	0,00	0,00	
<b>Indo-Européen, Germanique</b>	0,11	0,53	0,37	
<b>Indo-Européen, Hellénique</b>	0,00	0,29	0,71	
<b>Indo-Européen, Indo-Aryen</b>	0,33	0,67	0,00	
<b>Indo-Européen, Roman</b>	0,49	0,51	0,00	
<b>Indo-Européen, Slave</b>	0,04	0,51	0,45	<b>Légende</b>
<b>Japonique</b>	0,00	0,67	0,33	<b>1,00</b>
<b>Kx'a, #Hoan</b>	0,50	0,50	0,00	<b>0,90</b>
<b>Kx'a, Ju-Kung</b>	0,56	0,16	0,28	<b>0,80</b>
<b>Maya</b>	0,60	0,30	0,10	<b>0,70</b>
<b>Nadahup</b>	0,52	0,48	0,00	<b>0,60</b>
<b>Niger-Congo, Bantou</b>	0,57	0,34	0,09	<b>0,50</b>
<b>Niger-Congo, Edoïde</b>	0,78	0,22	0,00	<b>0,40</b>
				<b>0,30</b>

<b>Niger-Congo, Jukunoïde</b>	0,78	0,22	0,00	<b>0,20</b>
<b>Niger-Congo, Kwa</b>	0,60	0,36	0,04	<b>0,10</b>
<b>Niger-Congo, Senoufo</b>	0,78	0,22	0,00	<b>0,00</b>
<b>Nilo-Saharien, Saharien</b>	0,54	0,46	0,00	
<b>Ouralien, Fennique</b>	0,14	0,86	0,00	
<b>Ouralien, Ougrien</b>	0,00	0,45	0,55	
<b>Pama-Nyungan</b>	0,14	0,57	0,29	
<b>Sandawe</b>	0,50	0,50	0,00	
<b>Sino-Tibétain, Bodique</b>	0,22	0,33	0,44	
<b>Sino-Tibétain, Lolo-Birman</b>	0,44	0,56	0,00	
<b>Sino-Tibétain, Sinitique</b>	0,12	0,44	0,44	
<b>Tacana</b>	0,41	0,59	0,00	
<b>Tai-Kadai, Kadaï</b>	1,00	0,00	0,00	
<b>Tai-Kadai, Taï</b>	0,72	0,27	0,01	
<b>Tuu, !Ui</b>	0,60	0,40	0,00	

Tableau 4 : Proportion de Têtes, Relateurs et Satellites encodant de la Trajectoire pour chaque famille de langues

On observe dans ce tableau que l'encodage de la Trajectoire varie grandement entre les familles de langues. Certaines langues comme les langues viet-muong (austro-asiatique) encodent exclusivement, dans nos données, la trajectoire dans la Tête, comme l'indique la couleur rouge, signe que la totalité des formes encodant de la Trajectoire sont des Têtes. De même, les langues slaves (famille indo-européenne) privilégient plusieurs *loci* possibles tels que le Satellite et le Relateur. Ces multiples stratégies sont indiquées en jaune, signe que la moitié des formes encodant de la Trajectoire sont des Satellites, mais aussi des relateurs.

Cinq principales tendances sont observées dans nos données. Tout d'abord, l'encodage de la Trajectoire dans la **Tête** (H) est très présent dans les langues hmongiques (famille hmongmien), kadaï (famille taï-kadaï) et viet-muong (famille austro-asiatique). Les langues fenniques (famille ouralienne), quant à elles, encodent principalement la Trajectoire dans un **Relateur** (R).

Concernant les langues privilégiant plusieurs stratégies, les langues ju-kung (famille kx'a, anciennement khoïsan) encodent la Trajectoire dans la **Tête** et le **Satellite** (HS). On recense également des combinaisons **Tête** et **Relateur** (HR) dans les langues niger-congo, tchadiennes (famille afro-asiatique), sahariennes (famille nilo-saharienne), #hoan (famille kx'a, anciennement khoïsan) et sandawe (isolat, anciennement khoïsan), !ui (famille tuu,

anciennement khoïsan), nadahup, maya, Bahnariques (famille austro-asiatique), taïes (famille taï-kadaï), océaniques (famille austronésienne), indo-aryennes, romanes (famille indo-européenne), tacana, et lolo-birmanes (famille sino-tibétaine). Enfin, la dernière tendance observée est la stratégie **Relateur** et **Satellite** (RS), majoritaire en japonais et en coréen, ainsi que dans les langues slaves, germaniques, helléniques (famille indo-européenne), pama-nyunga, ougriennes (famille ouralienne), sinitiques et bodiques (famille sino-tibétaine). On note que les types S (satellite) et HRS (tête-relateur-satellite) ne sont dominants dans aucune famille de langues, bien qu’attestés dans différentes langues (respectivement exemples 5a et 5b).

5. a. Type S (satellite) : Anglais (Indo-Européen, Germanique)

He climbs **down**. (Fortis & Vittrant 2016)

b. Type HRS : Grec ancien (Indo-Européen, Hellénique ; Grèce)

**énth’**      hé:      g’      **eis-elthoûsa.**  
**to.there**    she      LNK      **to-go\_to.PART.AOR.NOM**  
 ‘Therein she entered’ (Fortis & Vittrant 2016)

## 5. CONCLUSION

Dans cette présentation de notre méthodologie, nous avons souhaité montrer comment utiliser des données linguistiques pour créer une ontologie des relations spatiales. Cette ontologie simplifie l’analyse de l’expression de la Trajectoire à partir de données provenant de langues différentes (typologiquement, génétiquement, et géographiquement). *Via* un système de codage, nous avons associé des formes linguistiques à leurs caractéristiques morphosyntaxiques (fonction syntaxique, partie du discours, catégorie grammaticale), et à leurs caractéristiques sémantiques (catégories conceptuelles du domaine spatial).

Type	Sous-type	Exemple de familles où le type est dominant
<b>H</b>		Langues hmong, kadaï, viet-muong
<b>R</b>		Langues fennique
<b>S</b>		<i>Non-attesté</i>
<b>HR</b>	<b>Plus de H</b>	Langues niger-congo, tchadiennes, nadahup, maya, taï
	<b>Plus de R</b>	Langues océaniques, romanes, tacana, lolo-birmanes



	<b>Autant de H et R</b>	Langues $\sharp$ hoan, sandawe
<b>HS</b>		Langues ju-kung
<b>RS</b>	<b>Plus de R</b>	Japonais, coréen, langues slaves, germaniques, pama-nyungan
	<b>Plus de S</b>	Langues ougriennes, hélléniques
	<b>Autant de R et S</b>	Langues sinitiques, bodiques
<b>HRS</b>		<i>Non-attesté</i>

Tableau 5 : Récapitulatif des *loci* de la Trajectoire

En utilisant cette ontologie, nous avons également établi une typologie préliminaire de l'encodage du composant de Trajectoire. Nous avons montré que cet encodage varie selon les familles de langues : les stratégies préférées par les langues pour encoder la Trajectoire (H : Tête, R : Relateur, S : Satellite, ou stratégies multiples) diffèrent selon leur appartenance à une famille. Ceci est résumé dans le tableau 5.

#### ABBREVIATIONS

**3** troisième personne, **ABL** ablatif, **ACC** accusatif, **AOR** aoriste, **CLF** classificateur, **CVB** coverbe, **GEN** génitif, **LNK** linker, **NOM** nominatif, **PART** participe, **PERL** perlatif, **PL** pluriel, **PST** passé

#### BIBLIOGRAPHIE

- Aske, Jon (1989). Path predicates in English and Spanish : A closer look. *Proceedings of the fifteenth annual meeting of the Berkeley Linguistic Society*, 1-14.
- Fagard, Benjamin, & Kopecka, Anetta (2021). Source/Goal (a)symmetry : A comparative study of German and Polish. In A. Kopecka & M. Vuillermet (Éds.), *Source-Goal (a)symmetries across languages*. Amsterdam: John Benjamins. 130-171.
- Farrar, Scott, & Langendoen, D. Terence (2003). A linguistic ontology for the semantic web. *GLOT international*, 7(3), 97-100.
- Fortis, Jean-Michel, & Vittrant, Alice (2011). L'organisation syntaxique de l'expression de la trajectoire : Vers une typologie de la construction. *Faits de Langues - Les Cahiers*, 3, 71-98.
- Fortis, Jean-Michel, & Vittrant, Alice (2016). On the morphosyntax of path-expressing constructions : Toward a typology. *STUF Language Typology and Universals*, 69(3), 341-374.

- Genesereth, Michael, & Nilsson, Nils (1987). *Logical Foundations of Artificial Intelligence*. Burlington: Morgan Kaufmann.
- Gruber, Thomas R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2), 199-220.
- Guarino, Nicola, Oberle, Daniel, & Staab, Steffen (2009). What is an ontology? In S. Staab & R. Studer (Éds.), *Handbook on ontologies*. Berlin: Springer-Verlag Berlin Heidelberg. 1-17.
- Guillaume, Antoine, & Koch, Harold (2021). Introduction : Associated Motion as a grammatical category in linguistic typology. In A. Guillaume & H. Koch (Éds.), *Associated motion*. Berlin, New York : De Gruyter. 3-30.
- Imbert, Caroline, Grinevald, Colette, & Söres, Anna (2011). Pour une catégorie de « satellite » de Trajectoire dans une approche fonctionnelle-typologique. *Faits de Langues - Les Cahiers*, 3, 99-116.
- Ishibashi, Miyuki, Kopecka, Anetta, & Vuillemer, Marine (2006). *Trajectoire : Matériel visuel pour élicitation des données linguistiques*. Fédération de Recherche en Typologie et Universaux Linguistiques.
- Matsumoto, Yo (2003). Typologies of Lexicalization Patterns and Event Integration : Clarifications and Reformulations. In S. Chiba (Éd.), *Empirical and theoretical investigations into language : A festschrift for Masaru Kajita*. Tokyo: Kaitakusha. 403-418.
- Matsumoto, Yo, Akita, Kimi, & Takahashi, Kiyoko (2017). The functional nature of deictic verbs and the coding patterns of Deixis : An experimental study in English, Japanese, and Thai. In I. Ibarretxe-Antunano (Éd.), *Motion and Space across Languages : Theory and applications*. Amsterdam: John Benjamins. 95-122.
- Morita, Takahiro (2009). *La catégorisation des verbes de déplacement en japonais et en français* [Thèse de doctorat]. EHES.
- Morita, Takahiro (2011). Intratypological Variations in Motion Events in Japanese and French. *CogniTextes. Revue de l'Association Française de Linguistique Cognitive*, 6, 6. [En ligne].
- Schalley, Andrea (2019). Ontologies and ontological methods in linguistics. *Language and Linguistics Compass*, 13(11), 1.
- Schalley, Andrea, Musgrave, Simon, & Haugh, Michael (2014). Accessing phonetic variation in spoken language corpora through non-standard orthography. *Australian Journal of Linguistics*, 34(1), 139-170.
- Seifen, Karl, Vittrant, Alice, Song, Jinke, & Bunkham, Nichuta (2021, 2/07). Path and Deixis as distinct concepts in Burmese, Thai, and Chinese. *NAMED 2021*, Paris, ENS.

- Slobin, Dan I., & Hoiting, Nini (1994). Reference to movement in spoken and signed languages : Typological considerations. *Proceedings of the Berkeley Linguistics Society*, 20, 487-505.
- Sowa, John F. (2003). *Ontology* [Blog]. <http://www.jfsowa.com/ontology/>
- Talmy, Leonard (1972). *Semantic Structures in English and Atsugewi* [Thèse de doctorat]. University of California.
- Talmy, Leonard (1985). Lexicalization patterns : Semantic structure in lexical form. In T. Shopen (Éd.), *Language Typology and Syntactic Description—Grammatical Categories and the Lexicon* (Vol. 3). Cambridge: Cambridge University Press. 57-149.
- Talmy, Leonard (2000). *Toward a Cognitive Semantics : Typology and Process in Concept Structuring* (Vol. 2). Cambridge: MIT Press.
- Vandeloise, Claude (1986). *L'Espace en français. Sémantique des propositions spatiale*. Paris : Seuil.
- Vuillermet, Marine (2012). *A grammar of Ese Ejja, a Takanan language of the Bolivian Amazon* [Thèse de Doctorat]. Université Lumière Lyon 2.
- Vuillermet, Marine, & Kopecka, Anetta (2019). Trajectoire : A methodological tool for eliciting Path of motion. *Methodological Tools for Linguistic Description and Typology, Language Documentation & Conservation Special Publication*, 16, 340-353.
- Wälchli, Bernhard (2001). A typology of displacement (with special reference to Latvian). *STUF - Language Typology and Universals*, 54, 298-323.