
**Gestion de ressources linguistiques —
Cadre d'annotation syntaxique
(SynAF) —**

**Partie 1:
Modèle syntaxique**

iTeh STANDARD PREVIEW
*Language resource management — Syntactic annotation framework
(SynAF) —
Part 1: Syntactic model*
(standards.iteh.ai)

ISO 24615-1:2014

<https://standards.iteh.ai/catalog/standards/sist/2aed3b97-4be7-483d-982a-5f00ca7e7328/iso-24615-1-2014>



iTeh STANDARD PREVIEW
(standards.iteh.ai)

[ISO 24615-1:2014](https://standards.iteh.ai/catalog/standards/sist/2aed3b97-4be7-483d-982a-5f00ca7e7328/iso-24615-1-2014)

<https://standards.iteh.ai/catalog/standards/sist/2aed3b97-4be7-483d-982a-5f00ca7e7328/iso-24615-1-2014>



DOCUMENT PROTÉGÉ PAR COPYRIGHT

© ISO 2014

Tous droits réservés. Sauf prescription différente ou nécessité dans le contexte de sa mise en œuvre, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie, ou la diffusion sur l'internet ou sur un intranet, sans autorisation écrite préalable. Une autorisation peut être demandée à l'ISO à l'adresse ci-après ou au comité membre de l'ISO dans le pays du demandeur.

ISO copyright office
Case postale 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Genève
Tél.: +41 22 749 01 11
Fax: +41 22 749 09 47
E-mail: copyright@iso.org
Web: www.iso.org

Publié en Suisse

Sommaire

Page

Avant-propos.....	iv
Introduction.....	v
1 Domaine d'application	1
2 Références normatives	1
3 Termes et définitions	1
4 Métamodèle SynAF	4
4.1 Introduction.....	4
4.2 Métamodèle SynAF.....	5
4.2.1 Vue d'ensemble.....	5
4.2.2 Classe SyntacticNode.....	5
4.2.3 Classe T_Node.....	5
4.2.4 Classe NT_Node.....	5
4.2.5 Classe SyntacticEdge.....	5
4.2.6 Classe Annotation.....	5
Annexe A (normative) Catégories de données pour SynAF	7
Annexe B (informative) Relation avec le cadre d'annotation linguistique	17
Bibliographie	19

iTeh STANDARD PREVIEW
(standards.iteh.ai)

[ISO 24615-1:2014](https://standards.iteh.ai/catalog/standards/sist/2aed3b97-4be7-483d-982a-5f00ca7e7328/iso-24615-1-2014)

<https://standards.iteh.ai/catalog/standards/sist/2aed3b97-4be7-483d-982a-5f00ca7e7328/iso-24615-1-2014>

Avant-propos

L'ISO (Organisation internationale de normalisation) est une fédération mondiale d'organismes nationaux de normalisation (comités membres de l'ISO). L'élaboration des Normes internationales est en général confiée aux comités techniques de l'ISO. Chaque comité membre intéressé par une étude a le droit de faire partie du comité technique créé à cet effet. Les organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO participent également aux travaux. L'ISO collabore étroitement avec la Commission électrotechnique internationale (IEC) en ce qui concerne la normalisation électrotechnique.

Les procédures utilisées pour élaborer le présent document et celles destinées à sa mise à jour sont décrites dans les Directives ISO/IEC, Partie 1. Il convient, en particulier de prendre note des différents critères d'approbation requis pour les différents types de documents ISO. Le présent document a été rédigé conformément aux règles de rédaction données dans les Directives ISO/IEC, Partie 2 (voir www.iso.org/directives).

L'attention est appelée sur le fait que certains des éléments du présent document peuvent faire l'objet de droits de propriété intellectuelle ou de droits analogues. L'ISO ne saurait être tenue pour responsable de ne pas avoir identifié de tels droits de propriété et averti de leur existence. Les détails concernant les références aux droits de propriété intellectuelle ou autres droits analogues identifiés lors de l'élaboration du document sont indiqués dans l'Introduction et/ou dans la liste des déclarations de brevets reçues par l'ISO (voir www.iso.org/brevets).

Les appellations commerciales éventuellement mentionnées dans le présent document sont données pour information, par souci de commodité, à l'intention des utilisateurs et ne sauraient constituer un engagement.

Pour une explication de la signification des termes et expressions spécifiques de l'ISO liés à l'évaluation de la conformité, ou pour toute information au sujet de l'adhésion de l'ISO aux principes de l'Organisation mondiale du commerce (OMC) concernant les obstacles techniques au commerce (OTC), voir le lien suivant: www.iso.org/iso/fr/avant-propos.html.

Le comité chargé de l'élaboration du présent document est l'ISO/TC 37, *Langage et terminologie*, sous-comité SC 4, *Gestion des ressources linguistiques*.

Cette première édition de l'ISO 24615-1 annule et remplace l'ISO 24615:2010 dont elle constitue une révision mineure.

L'ISO 24615, dans toutes ses parties, est conçue de manière coordonnée avec l'ISO 24612, *Gestion des ressources linguistiques – Cadre d'annotation linguistique (LAF)*, l'ISO 24613:2008, *Gestion des ressources linguistiques – Cadre d'annotation lexicale (LMF)*, et l'ISO 24611, *Gestion des ressources langagières – Cadre d'annotation morpho-syntaxique*.

L'ISO 24615 comporte les parties suivantes, sous le titre général *Gestion des ressources linguistiques – Cadre d'annotation syntaxique (SynAF)*:

— *Partie 1: Modèle syntaxique*

La partie suivante est en cours d'élaboration:

— *Part 2: Sérialisation XML (<Tiger2/>)*

Introduction

L'ISO 24615 se fonde sur de nombreux projets et activités de pré-normalisation qui ont été menés au cours des dernières années (voir Abeillé, 2001^[9]) en vue de constituer des modèles de référence et des formats pour la représentation de l'information syntaxique, soit sous forme de sortie d'analyseur syntaxique, soit sous forme d'annotations de ressources linguistiques (treebanks). Depuis plusieurs années l'initiative Penn Treebank constitue le standard *de facto* pour l'annotation syntaxique, mais des travaux plus récents, par exemple l'initiative Negra/Tiger (voir <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERCorpus/>) en Allemagne ou l'initiative ISST en Italie (voir Montemagni (2003)^[18]) démontre la viabilité d'un cadre plus cohérent qui tient compte à la fois des constituants hiérarchiques et des phénomènes de dépendance syntaxique dans l'annotation syntaxique.

Le projet eContent "LIRICS" est un projet majeur car il a rassemblé un groupe d'experts, initiateur du projet ISO 24615 (SynAF). Lors de la préparation de SynAF, ce groupe a confirmé que les initiatives actuelles partagent effectivement un modèle de données commun qui offre une bonne base de départ pour le métamodèle SynAF (voir l'étude "Evaluation of initiatives for morpho-syntactic and syntactic annotation" du projet européen LIRICS, disponible à l'adresse http://lirics.loria.fr/doc_pub/Del3_1_V2.pdf).

Cette partie de l'ISO 24615 propose un métamodèle pour l'annotation syntaxique ainsi qu'une liste de catégories de données pertinente pour l'annotation syntaxique. Les catégories de données sont disponibles sur le serveur ISOCat (<http://www.isocat.org/>) au sein du profil syntaxe (tel que défini dans l'ISO 12620:2009).

iTeh STANDARD PREVIEW (standards.iteh.ai)

ISO 24615-1:2014

<https://standards.iteh.ai/catalog/standards/sist/2aed3b97-4be7-483d-982a-5f00ca7e7328/iso-24615-1-2014>

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 24615-1:2014

<https://standards.iteh.ai/catalog/standards/sist/2aed3b97-4be7-483d-982a-5f00ca7e7328/iso-24615-1-2014>

Gestion de ressources linguistiques — Cadre d'annotation syntaxique (SynAF) —

Partie 1: Modèle syntaxique

1 Domaine d'application

Cette partie de l'ISO 24615 décrit le cadre d'annotation syntaxique (SynAF), modèle de haut niveau pour représenter l'annotation syntaxique des données linguistiques, dans le but d'offrir l'interopérabilité entre les ressources linguistiques ou les composants du traitement de la langue. La présente partie de l'ISO 24615 est complémentaire par rapport à l'ISO 24611 (MAF, cadre d'annotation morpho-syntaxique) et y est étroitement liée. Elle fournit un métamodèle pour les représentations syntaxiques, avec les catégories de données de référence destinées à représenter tout aussi bien des informations de constituance que des informations de dépendance, dans des phrases ou autres énonciations et segments du même ordre.

2 Références normatives

Les documents suivants sont indispensables à l'application de ce document. Pour les références datées, seule l'édition citée s'applique. Pour les références non datées, la dernière édition du document référencé s'applique (incluant ses éventuels amendements).

ISO 1087-1:2000, *Travaux terminologiques — Vocabulaire — Partie 1: Théorie et application*

ISO 12620:2009, *Terminologie et autres ressources langagières et ressources de contenu — Spécification de catégories de données et gestion d'un registre de catégories de données pour les ressources langagières*

ISO 24611:2012, *Gestion des ressources langagières — Cadre d'annotation morphosyntaxique (MAF)*

3 Termes et définitions

Pour les besoins du présent document, les termes et définitions figurant dans l'ISO 1087-1:2000, l'ISO 12620:2009, l'ISO 24611:2012 ainsi que ce les suivants s'appliquent.

3.1 circonstant

élément non essentiel associé à un verbe en opposition avec les *arguments syntaxiques* (3.19)

Note 1 à l'article: Les adverbes sont des circonstants possibles pour une phrase.

3.2 bloc bloc syntaxique

constituant non récursif (3.4)

Note 1 à l'article: Le terme «bloc» est la traduction en français du terme anglais «chunk»

3.3

proposition

groupe de syntagmes (3.14), qui comprend habituellement un prédicat

Note 1 à l'article: Une proposition peut être soit une *proposition principale* (3.10) soit une *proposition subordonnée* (3.17). Dans les langues qui distinguent la finitude, les propositions dont le prédicat est un verbe peuvent être soit finies soit non-finies, en fonction de la forme du verbe. Une proposition principale peut constituer à elle seule une *phrase* complète (3.15). Dans le modèle SynAF, une proposition est un cas particulier de *constituant* (3.4).

3.4

constituant

groupement syntaxique de mots [en *syntagmes* (3.14)], de syntagmes [en *propositions* (3.3) ou d'autres *syntagmes*] ou de *propositions* [en une *phrase* (3.15)], sur la base de propriétés structurelles (ou hiérarchiques)

3.5

dépendance

relation de dépendance

relation syntaxique entre *mots-formes* (3.24) ou *constituants* (3.4) sur la base des *fonctions grammaticales* (3.7) que les *constituants* assument les uns par rapport aux autres

3.6

arc syntaxique

arc

triplet avec un *nœud* source (3.12), un *nœud* cible, et des *annotations* optionnelles (3.9)

Note 1 à l'article: Les *nœuds non-terminaux* (3.13) ont un arc syntaxique sortant de constituance.

3.7

fonction grammaticale

rôle grammatical d'un *mot-forme* (3.24) ou d'un *constituant* (3.4) dans son environnement syntaxique enchaissant

<https://standards.iteh.ai/catalog/standards/sist/2aed3b97-4be7-483d-982a-5f00ca7e7328/iso-24615-1-2014>

Note 1 à l'article: par exemple, un syntagme nominal (SN) peut jouer le rôle du sujet dans une *phrase* (3.15), ou un nom celui du sujet dépendant d'un verbe dans un graphe de dépendance. Il y a un lien grammatical entre le sujet – SN et le verbe principal d'une phrase. Toutes les relations grammaticales (sujet – prédicat, tête – modificateur, etc.) sont subsumées sous le concept de *relation de dépendance* (3.5), que cela soit entre nœuds terminaux ou nœuds non-terminaux.

3.8

tête syntaxique

tête

partie d'un *constituant* (3.4) qui détermine sa distribution (l'environnement syntaxique dans lequel le constituant peut apparaître) et ses propriétés grammaticales (ex: si le genre grammatical de la tête est féminin, alors le genre de la totalité du constituant sera féminin)

Note 1 à l'article: La tête d'un constituant ne peut habituellement pas être négligée.

3.9

annotation linguistique

annotation

paire attribut-valeur dénotant une propriété linguistique d'un segment linguistique

3.10

proposition principale

proposition (3.3) qui peut fonctionner toute seule comme une *phrase* complète (3.15)

Note 1 à l'article: Dans les langues distinguant la finitude, la proposition principale est habituellement finie. Exemple: *Le train est en retard.*

3.11 modificateur

partie d'un *constituant* (3.4) qui attribue une propriété à la *tête* (3.8) dudit constituant

Note 1 à l'article: Un modificateur peut être situé avant ou après la tête du syntagme (3.14) (pré-modificateur ou post-modificateur). Dans un constituant, les modificateurs sont optionnels.

3.12 nœud nœud syntaxique

mot-forme (3.24) ou *constituant* (3.4) considéré comme un composant syntaxique élémentaire d'une analyse syntaxique

3.13 nœud non terminal

nœud syntaxique (3.12) qui n'est pas un *mot-forme* (3.24)

Note 1 à l'article: Un nœud non terminal possède un *arc* sortant de constituance (3.6).

3.14 syntagme

groupe de *mots-formes* (3.24) (comportant habituellement un ou plusieurs mots) qui peut remplir une *fonction grammaticale* (3.7), dans une *proposition* (3.3), par exemple

Note 1 à l'article: Les syntagmes vides sont autorisés (pronoms non-réalisés, quelquefois marqués comme «pro» et ayant le rôle de sujets dans des propositions). Un syntagme est typiquement nommé d'après sa *tête* (3.8), par exemple les syntagmes nominaux, les syntagmes verbaux, les syntagmes adjectivaux, les syntagmes adverbiaux et les syntagmes prépositionnels. De manière informelle, les syntagmes sont décrits comme des «mots gonflés», du fait que les éléments qui le constituent et qui sont ajoutés à la tête syntaxique forment et spécifient la référence de celle-ci. Dans notre modèle, le syntagme est un cas particulier de *constituant* (3.4).

3.15 phrase

groupe lié de *mots-formes* (3.24) contenant une prédication, exprimant habituellement une pensée complète et formant l'unité de base de la structure de discours

Note 1 à l'article: Une phrase consiste en une ou plusieurs *propositions* (3.3). Quand l'objectif est de décrire le discours, il est commun de parler d'énonciations plutôt que de phrases.

3.16 empan

paire de points (p1, p2), pour lesquels $p1 \leq p2$, identifiant les segments du document faisant l'objet de l'*annotation* (3.9)

Note 1 à l'article: Un empan multiple est une séquence d'empans dans laquelle le point final de chaque empan est inférieur ou égal au point de départ de l'empan suivant.

3.17 proposition subordonnée

proposition qui remplit une *fonction grammaticale* (3.7) dans un *syntagme* (3.14) [par exemple, une *proposition* relative (3.3) modifiant la *tête* (3.8) nominale d'un syntagme nominal] ou dans une autre proposition

Note 1 à l'article: Généralement, une proposition subordonnée ne fonctionne pas seule comme une phrase, elle constitue une partie d'une phrase plus grande.

3.18 cadre de sous-catégorisation

ensemble de restrictions indiquant les propriétés des *arguments syntaxiques* (3.19) qui peuvent ou doivent apparaître avec un verbe

EXEMPLE Alfred (/argument syntaxique/) lit un livre (/argument syntaxique/) aujourd'hui (/circonstant/).

Note 1 à l'article: Le sujet, l'objet indirect et l'objet direct sont des *fonctions grammaticales* (3.7) sous-catégorisées au sein d'une phrase; elles sont dépendantes du verbe (c'est-à-dire qu'elles peuvent apparaître dans des cadres de sous-catégorisation).

3.19 argument syntaxique

élément fonctionnellement essentiel qui est requis et dont l'interprétation est donnée par la tête de son *syntagme* (3.14) ou le *nœud* (3.12) dont il dépend (par ex, l'argument nominal d'un syntagme prépositionnel ou d'un verbe)

Note 1 à l'article: Pour les verbes et les syntagmes verbaux, les arguments identifient les participants au processus référencé par le verbe. Dans certains cadres, les arguments syntaxiques sont appelés compléments.

3.20 graphe syntaxique graphe

ensemble connecté de *nœuds syntaxiques* (3.12) et d'*arcs* (3.6)

3.21 arbre syntaxique

graphe syntaxique (3.20) dans lequel chaque nœud n'a qu'un seul parent

3.22 syntaxe

façon dont les mots-formes (3.24) sont interconnectés et/ou groupés ensemble pour former des syntagmes, et ainsi fixer les relations qui existent entre lesdits éléments

3.23 nœud terminal

nœud syntaxique (3.12) qui est un *mot-forme* (3.24) simple ou un élément vide impliqué dans une relation syntaxique

3.24 mot-forme

entité contiguë ou non-contiguë dans une séquence de discours ou de texte, identifiée comme un item lexical autonome

ITEH STANDARD PREVIEW

(standards.iteh.ai)

ISO 24615-1:2014

<https://standards.iteh.ai/catalog/standards/sist/2aed3b97-4be7-483d-982a-5f00ca7e7328/iso-24615-1-2014>

4 Métamodèle SynAF

4.1 Introduction

Les annotations syntaxiques ont au moins deux fonctions dans le traitement du langage:

- a) représenter la constituance linguistique, comme dans les syntagmes nominaux (SN), en décrivant une séquence structurée d'items annotés en morpho-syntaxe (incluant des éléments vides ou des traces engendrées par des mouvements au niveau de la constituance), ainsi que les constituants construits à partir d'éléments non-contigus, et
- b) représenter les relations de dépendance, comme les relations tête-modificateur, et incluant les relations entre catégories de même nature (comme les relations tête-tête entre noms dans les appositions, ou les coordinations nominales dans certains formalismes). L'information de dépendance peut exister entre items annotés en morpho-syntaxe au sein d'un syntagme (un adjectif est le modificateur d'une tête nominale au sein d'un SN) ou bien décrit une relation spécifique entre constituants syntaxiques au niveau de la proposition ou de la phrase (c'est-à-dire, un SN «sujet» du verbe principal de la proposition ou de la phrase). La relation de dépendance peut aussi être associée à des éléments vides (par exemple, l'élément pro dans les langues romanes qui assume une fonction grammaticale).

En conséquence, les annotations syntaxiques doivent être conformes à une stratégie d'annotation multicouches mettant en corrélation les annotations syntaxiques relatives à la constituance et à la dépendance, comme spécifié dans le métamodèle SynAF.

4.2 Métamodèle SynAF

4.2.1 Vue d'ensemble

Le métamodèle SynAF est représenté comme un ensemble de classes UML complété par des paires attribut-valeur, qui représentent les catégories de données syntaxiques. La description textuelle SynAF spécifie de manière plus complète les classes SynAF, les relations et extensions qui peuvent être incluses dans le diagramme UML. Les développeurs doivent définir une sélection de catégories de données (DCS) comme spécifié dans les procédures de sélection des catégories de données pour SynAF (voir [Figure 1](#)). Les catégories de données fournies dans l'[Annexe A](#) doivent être utilisées pour la représentation des annotations syntaxiques.

4.2.2 Classe SyntacticNode

La classe *SyntacticNode* est une classe générique subsumant à la fois la classe des nœuds terminaux et la classe des nœuds non-terminaux. Les nœuds syntaxiques peuvent être impliqués dans autant de relations syntaxiques que nécessaires (voir [3.6 arc syntaxiques](#)).

4.2.3 Classe T_Node

La classe *T_Node* représente les nœuds terminaux d'un arbre syntaxique, constitué de mots-formes annotés en morpho-syntaxe, ainsi que d'éléments vides, s'il y a lieu. Les nœuds T_Nodes sont définis sur un ou plusieurs empan (des empan multiples peuvent servir pour les constituants discontinus). Les T_Nodes sont annotés avec des catégories syntaxiques valides au niveau du mot.

4.2.4 Classe NT_Node

La classe *NT_Node* représente les nœuds non-terminaux de l'arbre syntaxique. Les arbres syntaxiques sont principalement constitués de T_Nodes et de NT_Nodes, avec des éléments vides, s'il y a lieu. Les T_Nodes font référence à un empan. De ce fait, en vertu des représentations de l'arbre syntaxique, des empan peuvent aussi être déduits pour les NT_Nodes. Les NT_Nodes sont annotés avec des catégories syntaxiques valides au niveau syntagmatique et supérieur (niveau propositionnel ou phrastique).

4.2.5 Classe SyntacticEdge

La classe *SyntacticEdge* représente une relation entre nœuds syntaxiques (à la fois nœuds terminaux et non-terminaux). Par exemple, la relation de dépendance est binaire, composée d'une paire de nœuds source et cible, avec une ou plusieurs annotations. En particulier, un arc syntaxique peut être annoté par un */syntacticEdgeType/* (voir l'[Annexe A](#)), dont le domaine conceptuel peut être un */primarySyntacticEdge/* ou */secondarySyntacticEdge/*, sans que cela soit limitatif.

4.2.6 Classe Annotation

La classe *Annotation* représente l'application de l'information syntaxique aux données annotées SynAF, ainsi que l'application de l'information morpho-syntaxique aux données annotées MAF (voir [Figure 1](#)).