



CATHERINE ROUSSEY, MARINE COURTIN, ROBERT BOSSY, STEPHAN BERNARD,
XAVIER DELPUECH, LIONEL DELBAC

Description sémantique des stades de développement phénologique des plantes, et
extraction de nouveaux labels

Volume 6, n° 1-2 (2025), p. 157-178.

<https://doi.org/10.5802/roia.97>

© Les auteurs, 2025.



Cet article est diffusé sous la licence

CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.

<http://creativecommons.org/licenses/by/4.0/>



*La Revue Ouverte d'Intelligence Artificielle est membre du
Centre Mersenne pour l'édition scientifique ouverte*

www.centre-mersenne.org

e-ISSN : 2967-9672

Description sémantique des stades de développement phénologique des plantes, et extraction de nouveaux labels

Catherine Roussey^a, Marine Courtin^b, Robert Bossy^b,
Stephan Bernard^c, Xavier Delpuech^d, Lionel Delbac^{® e}

^a UMR MISTEA, Université de Montpellier, INRAE, Institut Agro, Montpellier, France
E-mail : catherine.roussey@inrae.fr

^b MaIAGE, INRAE, Jouy-en-Josas, France
E-mail : marine.courtin@inrae.fr, robert.bossy@inrae.fr

^c UR LISC, Université Clermont Auvergne, INRAE, Aubière, FRANCE
E-mail : stephan.bernard@inrae.fr

^d Institut français de la vigne et du vin, Pôle Rhône-Méditerranée, Montpellier, France
E-mail : xavier.delpuech@vignevin.com

^e UMR SAVE, INRAE, Bordeaux Sciences Agro, Université de Bordeaux, Villenave d'Ornon, France
E-mail : lionel.delbac@inrae.fr.

RÉSUMÉ. — Le projet ANR « Des Données aux Connaissances en Agronomie et Biodiversité » (D2KAB) met à disposition une archive de bulletins agricoles publiée sur le Web. Pour annoter les bulletins à l'aide des stades de développement des plantes, nous avons besoin d'une nouvelle ressource sémantique. Plusieurs échelles phénologiques existent, chacune de ces échelles définit un ensemble spécifique de stades. L'Institut Français de la Vigne et du vin (IFV) a aligné plusieurs échelles existantes associées à la vigne. Nous présentons ici la création d'un cadre méthodologique pour la description sémantique des stades de développement des plantes fondé sur l'ontologie PPDO (*BBCH-based Plant Phenological Description Ontology*). Cette ontologie a été peuplée pour construire plusieurs échelles phénologiques associées à différentes cultures : vigne, céréales, légumes, arboriculture. De plus, une expérimentation d'annotation des bulletins viticoles à partir des échelles de la vigne a été menée pour évaluer l'utilisation des stades.⁽¹⁾

MOTS-CLÉS. — Ontologie, ingénierie ontologique, Web de données, Données liées, Stades phénologiques /de développement, Agriculture numérique, Vigne, BBCH, annotation, bulletin agricole, bulletin de santé du végétal.

1. INTRODUCTION

L'agronomie et l'agriculture sont confrontées à plusieurs défis sociétaux, économiques et environnementaux majeurs, qu'une approche sémantique de la science

des données aidera à relever. Le projet ANR *Des Données aux Connaissances en Agronomie et Biodiversité* (D2KAB)⁽¹⁾ illustre comment la science des données sémantiques contribue au développement d'applications agricoles innovantes. L'objectif de D2KAB est de créer un cadre pour transformer les données d'agronomie et de biodiversité en connaissances décrites sémantiquement, interopérables, exploitables et ouvertes. Pour construire un tel cadre, nous nous appuyons sur des ressources sémantiques (par exemple, des terminologies, des vocabulaires ou des ontologies) pour décrire nos données et les publier en tant que données ouvertes liées [8]. Nous utilisons le portail web AgroPortal⁽²⁾ [19] pour trouver, publier et partager des ressources sémantiques puis nous exploitons ces ressources sémantiques dans des applications dédiées à l'agriculture ou l'environnement.

L'un des scénarios agricoles de D2KAB consiste à construire un navigateur web augmenté pour les bulletins officiels d'alertes agricoles français, appelés *Bulletins de Santé du Végétal* (BSV). Les BSV assurent la diffusion des observations sanitaires réalisées sur les réseaux de surveillance des bioagresseurs pour une culture donnée [30]. Le prototype interrogera une archive de BSV disponible sous forme de fichiers PDF. Chaque bulletin est annoté sémantiquement. Ces annotations sont produites à partir de techniques de traitement automatique de la langue appliquées sur les contenus textuels des BSV. Les annotations sont décrites en utilisant l'ontologie Web Annotation Data Model et publiées sur le Web de données liées [35].

Des mentions de stades phénologiques sont toujours présentes dans les BSV, afin d'identifier les périodes de sensibilité d'une culture aux facteurs abiotiques (par exemple, le gel printanier) ou à des facteurs biotiques (par exemple, la maladie du mildiou). Afin d'annoter les stades phénologiques des cultures, nous avons besoin d'une ressource sémantique spécifique pour représenter les stades, l'échelle phénologique à laquelle ils appartiennent et les cultures auxquelles ils s'appliquent. Comme il existe plusieurs échelles phénologiques par culture, nous avons aussi besoin d'aligner les stades des échelles différentes concernant la même culture. Il est à noter que la plupart des échelles phénologiques sont dédiées à un ensemble limité de culture voir même une seule culture. Une des échelles les plus connues et qui s'applique à plusieurs cultures est l'échelle BBCH [23]. Malheureusement, ces échelles ne sont pas publiées sur le Web de données liées, ni disponibles dans un format du Web sémantique (e.g., SKOS, RDF-S ou OWL), empêchant leur utilisation dans différentes applications agricoles. Notre corpus de BSV montre que différentes échelles relatives à une même culture ont été utilisées pour observer les stades phénologiques. Par exemple, dans un BSV lié à la vigne, on trouve la phrase : *bourgeon dans le coton (stade 03 ou B ou BBCH 05) dans les secteurs tardifs*, qui est à la fois une référence à un « label » (« bourgeon dans le coton ») et la codification de plusieurs échelles : (i) 03 est un code de l'échelle Eichhorn–Lorenz [20]; (ii) B est un code de l'échelle de Baggiolini [6]; (iii) et BBCH 05 est un code de l'échelle BBCH [23]. Ces codes devront être identifiés automatiquement dans un texte pour être associés à un identifiant pérenne représentant le stade

⁽¹⁾www.d2kab.org

⁽²⁾<http://agroportal.lirmm.fr>

d'une échelle phénologique. De plus, il semble que les labels les plus fréquemment utilisés pour la vigne soient ceux d'une échelle produite par l'Institut Français de la Vigne et du vin (IFV). Malheureusement cette échelle propose uniquement des labels. D'après notre expérience sur l'usage des BSV, il est nécessaire de construire un ensemble de graphes de connaissances alignés pour décrire sémantiquement les stades des différentes échelles.

Cet article est une traduction et extension d'un article précédent [31]; nous présentons la finalisation de notre travail de transformation en RDF de deux échelles génériques BBCH, de cinq échelles de la vigne (dont l'échelle individuelle BBCH de la vigne), et de sept échelles BBCH individuelles (non présentées dans l'article précédent).

Nous décrirons plus particulièrement notre travail d'alignement entre les échelles de la vigne. Nous présenterons la dernière version de l'ontologie intitulée *BBCH-based Plant Phenological Description Ontology (PPDO)*. Cette ontologie est le support d'un cadre méthodologique pour encoder sémantiquement toute échelle phénologique. Cette ontologie décrit des sous ensemble de stades sous forme de classes OWL. Puis, chaque stade d'une échelle phénologique est représentée par une instance des classes précédentes.

La suite de l'article est organisé comme suit : la section 2 introduit la phénologie et présente les échelles que nous avons transformées ; la section 3 passe en revue les thésaurus/ontologies existants pour décrire les stades phénologiques ; la section 4 illustre la méthodologie utilisée pour construire l'ontologie PPDO et la peupler avec les stades des échelles phénologiques ; la section 5 présente une expérimentation d'annotation des bulletins viticoles avec les échelles de la vigne ; la section 6 discute des avantages et des inconvénients de la publication des échelles phénologiques sous forme de données ouvertes liées.

2. ÉCHELLES PHÉNOLOGIQUES

La phénologie des plantes est l'étude du développement saisonnier des végétaux déterminé par l'influence des variations du climat. Les événements périodiques sont par exemple l'apparition des fleurs, le changement de couleurs des feuilles, etc. Un stade phénologique ou stade de développement des plantes caractérise une phase de développement de la plante pendant son cycle de vie. Par exemple, le stade floraison est observé quand au moins 50 % des fleurs sont épanouies. L'étude phénologique d'une plante consiste à observer à quelle date les stades apparaissent. Les variations temporelles sur l'apparition d'un stade peuvent dépendre du climat ou d'autres facteurs comme la variété de la plante.

L'échelle phénologique BBCH améliorée (*Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie*) propose une codification homogène des stades de développement communs à différentes espèces végétales cultivées. BBCH décrit plusieurs ensembles de stades de développement : l'échelle générale et des échelles spécifiques par culture (dites « échelles individuelles ») [23]. L'échelle générale forme le

cadre dans lequel les échelles individuelles sont développées. Elle peut également être utilisée pour les cultures n'ayant pas fait l'objet de description d'échelle individuelle : *« Les différentes phases du développement d'une plante sont divisées en dix stades principaux numérotés de 0 à 9. (...) Pour définir précisément les stades d'application ou d'évaluation, il n'est pas conseillé d'utiliser seulement les stades principaux, car ceux-ci recouvrent une durée importante dans le développement de la plante. On utilise les stades de développement secondaires pour déterminer un moment exact du développement. Par rapport aux stades principaux, les stades secondaires définissent des stades limités dans le temps. Ils sont donnés par des valeurs de 0 à 9, qui se suivent à l'intérieur d'un stade principal. On obtient ainsi un code à deux chiffres composé par le stade principal et le stade secondaire. »* [23]. BBCH est considéré comme une référence pour décrire les stades et la monographie est disponible en anglais, français, allemand et espagnol [21, 22, 23, 24].

Pour la vigne, la première échelle phénologique a été proposée par Baggiolini [5] en 1952. Elle comportait initialement 10 stades qui ont été complétés en 1993 par Baillod et Baggiolini [6] pour atteindre 16 stades. En 1977, Eichhorn et Lorenz ont proposé une échelle (EL) plus détaillée de 24 stades pour la vigne. En 1992, une échelle universelle connue sous le nom de BBCH pour les plantes cultivées mono et dicotylédones a été proposée par Hack et al. [16], puis adaptée à la vigne en 1995 par Lorenz et al. [20]. L'échelle individuelle BBCH de la vigne permet la comparaison avec d'autres espèces, aussi bien annuelles que pérennes, et s'est progressivement imposée comme une échelle de référence dans la communauté scientifique.

En 1995, Coombe [12] a proposé un alignement entre les différentes échelles existantes de la vigne tout en proposant des modifications de l'échelle EL. Cependant, dans la communauté technique agricole française, différentes échelles phénologiques sont parfois utilisées selon les habitudes des techniciens. L'IFV a donc produit une fiche technique [17] pour synthétiser les stades de développement de la vigne les plus utiles et les plus pertinents. Les stades sélectionnés ont été définis à l'aide d'expressions françaises spécifiques et illustrés par des images représentatives. Leurs correspondances avec les trois échelles phénologiques (échelle individuelle BBCH, EL et Baggiolini) ont été indiquées. Cette échelle, qui peut être vue comme une sélection des stades BBCH, porte ici le nom de IFV-Labels.

L'échelle IFV-Epicure complète celle d'EL et propose d'objectiver, par des éléments de mesure (pourcentage ou métrique), certaines définitions de stades trop imprécises. Le développement de la base de données Epicure au début des années 2000 a conduit l'IFV de Bordeaux à standardiser l'observation de stades phénologiques recueillis sur le terrain par les viticulteurs et techniciens. L'échelle IFV-Epicure reprend donc la numérotation proposée par EL, car c'est la plus utilisée en pratique sur le vignoble. Elle propose une numérotation continue du stade repos d'hiver (1) à la fin de la chute des feuilles (47) en occupant tous les numéros de stades laissés vacants par EL. Ce faisant, IFV-Epicure propose un échelonnage plus régulier et affine l'observation de certains stades trop imprécis par adjonction de critères mesurables : par exemple, le

stade EL 31 (ou K de Baggiolini) défini par « grains de pois » reste le stade 31 d'IFV-Epicure mais est complété par une évaluation du diamètre des baies (4 à 5 mm). Nos travaux ont permis de traduire et d'aligner les cinq échelles de la vigne :

- l'échelle individuelle BBCH de la vigne [23];
- l'échelle Eichhorn–Lorenz [20] qui ne concerne que la vigne;
- l'échelle Baggiolini de la vigne [6];
- l'échelle produite par IFV qui est une sélection de l'échelle individuelle BBCH de la vigne (intitulée IFV-labels);
- l'échelle produite par IFV qui est un enrichissement de l'échelle Eichhorn–Lorenz (intitulée IFV-Epicure).

Nous avons aussi traduit l'échelle générale BBCH et créé une autre échelle générique BBCH dite échelle globale qui a été construite artificiellement en agrégeant tous les stades des échelles BBCH existantes. Ces deux échelles génériques sont encodées dans un graphe prefixé *ppd-gen*.

Même s'il existe d'autres échelles pour d'autres cultures, comme Zadoks (1974) [36] pour le blé, il n'y a manifestement pas de liste de stades phénologiques universellement reconnue; cependant, en raison de sa large adoption, la codification proposée dans les échelles BBCH peut être vue comme un langage pivot pour décrire les stades de développement spécifiques aux cultures. En d'autres termes, les graphes RDF décrivant les stades spécifiques d'une culture devront être alignés avec une des échelles BBCH (soit *via* une instanciation de classes, soit *via* des correspondances explicites entre l'échelle BBCH spécifique à la culture). C'est pour cette raison que nous avons publié sept échelles individuelles BBCH supplémentaires :

- l'échelle individuelle BBCH des céréales qui s'applique au blé, à l'orge, à l'avoine et au seigle (prefix *ppd-cr*);
- l'échelle individuelle BBCH des légumes feuilles formant des pommes qui s'applique aux choux, aux choux chinois, aux laitues pommées et aux endives (prefix *ppd-lvfh*);
- l'échelle individuelle BBCH des légumes racines ou tubercules qui s'applique aux carottes, aux céleris, aux choux-raves, aux chicorées, aux radis, aux rutabagas et aux scorsonères (prefix *ppd-rsv*);
- l'échelle individuelle BBCH des légumes solanacées qui s'applique aux tomates, aux aubergines et aux poivrons (prefix *ppd-sol*);
- l'échelle individuelle BBCH des fruits à noyaux qui s'applique aux cerisiers, aux pruniers, aux pêchers et aux abricotiers (prefix *ppd-stf*);
- l'échelle individuelle BBCH des oliviers (prefix *ppd-olt*);
- l'échelle individuelle BBCH des fruits à pépins qui s'applique aux pommiers et aux poiriers (prefix *ppd-pof*).

3. RESSOURCES SÉMANTIQUES DÉCRIVANT LES STADES PHÉNOLOGIQUES

Aucune des échelles mentionnées précédemment n'est publiée sur le Web de données liées par leurs producteurs. Cependant, certains schéma RDF ou ontologies OWL existants mentionnent des stades de développement. La recherche du mot clé «BBCH» sur AgroPortal nous retourne plusieurs ressources sémantiques : la Crop Ontology (CO) et ses extensions *Woody Plant Ontology* (CO_357), *Brassica Ontology* (CO_348), *Wheat Plant Anatomy and Development Ontology* (CO_212), la Plant Ontology (PO) et SOY (une ontologie décrivant les traits spécifiques du soja). Il est à noter que le thesaurus ANAEE [11] sur la biodiversité possède un concept pour représenter les échelles BBCH. Nous avons également identifié d'autres ressources sémantiques pouvant contenir des description de stades phénologiques, notamment le thesaurus AGROVOC [29] et la Plant Phenology Ontology (PPO) [33].

3.1. AGROVOC

Le thesaurus AGROVOC est publié par l'Organisation des Nations Unies pour l'Alimentation et l'Agriculture (FAO) [29]. Il est édité par une communauté mondiale d'experts et couvre tous les domaines d'intérêt de la FAO, y compris l'agriculture, la sylviculture, la pêche, l'alimentation et les domaines connexes. Il est disponible en 29 langues, avec une moyenne de 35 000 termes par langue et développé en SKOS-XL. La force de ce thesaurus est sa couverture lexicale multilingue. Il est donc souvent utilisé pour annoter ou indexer des documents ou des images relatifs au domaine de l'agriculture.

AGROVOC contient certains stades phénologiques. L'instance de skos:Concept intitulé «stades de développement végétal» est la racine d'une hiérarchie de 38 stades. Par exemple, l'URI http://aims.fao.org/aos/agrovoc/c_2992 identifie le stade «floraison». Notons que la référence à une échelle existante n'est pas indiquée – c'est-à-dire qu'aucun texte associé à l'instance ne fait référence à une échelle en particulier comme BBCH et qu'aucun lien externe ne pointe vers une information de ce type. Il en est de même pour la culture, aucune information ne permet de dire à quelle culture le stade peut s'appliquer. Pour ces raisons, AGROVOC ne correspond pas à notre objectif. Il pourra en revanche être une source externe à laquelle s'aligner pour plus d'interopérabilité.

3.2. LA PLANT ONTOLOGY ET LA CROP ONTOLOGY

La *Plant Ontology* (PO) [13] et la *Crop Ontology* (CO) [32], pour les plantes cultivés, sont deux résultats du projet Planteome [18]. L'objectif était de développer des ontologies de référence en génomique et en phénomique végétale. CO est le regroupement de plusieurs dictionnaires de traits observables spécifiques à une culture, tous liés à PO ainsi qu'à la *Plant Trait Ontology* (TO). L'un des dictionnaires de traits spécifiques à la vigne est l'ontologie *Vitis* (CO_356),⁽³⁾ mais elle ne contient aucune mention de stade phénologique, au contraire d'autres extensions de CO comme

⁽³⁾http://agroportal.lirmm.fr/ontologies/CO_356

Brassica Ontology (CO_348). CO_348 contient 10 classes représentant des traits phénologiques associés à des labels de BBCH, comme la classe intitulée « Beginning of flowering time (BBCH 61) ». Une des échelles individuelles BBCH relative au genre brassica (dont les choux font partie) est intitulée « légumes feuilles formant des pommes » et contient plus de 10 stades secondaires. Donc CO_348 ne couvre pas l'ensemble des besoins. PO couvre toutes les espèces cultivées ; elle contient des termes normalisés, des définitions et des relations décrivant l'anatomie, la morphologie et les stades [34]. Dans PO, un stade phénologique est représentée par une classe. Un stade correspond à un intervalle de temps pendant le cycle de croissance d'un élément végétal. Ainsi, un stade bref est défini comme une sous-classe d'un stade plus long englobant ce stade bref. Un stade est lié à une entité anatomique végétale à l'aide de la propriété objet *has_participant*. Certaines références à des échelles phénologiques ont été trouvées dans les valeurs des propriétés d'annotation *has_related_synonym* de PO : 49 pour l'échelle BBCH, 38 pour l'échelle Zadoks. Par exemple, l'URI http://purl.obolibrary.org/obo/PO_0007086 correspond au stade « cinq nœuds visibles ». Le code de ce stade dans l'échelle BBCH est BBCH 35. Comme la plupart des ontologies de la communauté OBO, PO a pour base l'ontologie « fondatrice » *Basic Formal Ontology* (BFO) [3] et réutilise des propriétés définies dans *Relation Ontology* (RO). Nous avons trouvé dans PO des propriétés (object properties) intéressantes. Les propriétés *precedes* et *preceded_by* indiquent que la fin d'un stade se produit avant le début d'un autre. Ces propriétés définissent des restrictions de domaine et de co-domaine liées aux classes de BFO. Ainsi, réutiliser ces propriétés signifie s'appuyer sur les classes de BFO. PO a été conçue pour annoter les données génomiques et phénotypiques. Cette ontologie est plus détaillée que l'échelle BBCH qui elle a été conçue pour les pratiques agricoles. Par conséquent, nous avons jugé PO non pertinente pour notre cas d'usage.

3.3. PLANT PHENOLOGY ONTOLOGY (PPO)

La *Plant Phenology Ontology* (PPO) pour la phénologie végétale fournit un vocabulaire standardisé utile pour l'intégration de données phénologiques hétérogènes du monde entier [33]. Elle s'appuie sur PO ainsi que sur *Phenotype and Trait Ontology* (PATO) pour décrire les phénotypes et les traits. La phénologie n'est pas l'observation des stades phénologiques à proprement parler ; mais l'observation des traits physiques des plantes, comme la présence de feuilles ou de fruits. Ainsi, PPO définit les traits observables et ne se concentre pas sur l'observation de début ou de la fin d'un stade. PPO permet de décrire précisément la présence ou l'absence de certaines entités anatomiques végétales modélisées comme des traits.

Pour conclure, parmi les ressources sémantiques identifiées, nous n'avons trouvé aucune solution à adopter directement pour notre cas d'usage. Bien que BBCH soit souvent référencé, ce qui confirme son rôle de langage pivot, nous n'avons pas trouvé de ressource sémantique qui s'appuierait sur BBCH et offrirait un moyen d'encoder sémantiquement les stades phénologique de toutes les cultures. Par conséquent, nous avons décidé de proposer un nouveau cadre méthodologique pour décrire les échelles

phénologiques : ce cadre contiendra une ontologie OWL déclarant un ensemble de classes spécialisant le modèle SKOS ainsi que plusieurs graphes de connaissances SKOS. Un graphe SKOS déclare l'ensemble des stades d'une échelle spécifique à une culture – dans le cas de la vigne nous aurons donc encodé et aligné les échelles suivantes : échelle individuelle BBCH de la vigne, les échelles EL et Baggiolini, les échelles produites par IFV intitulée IFV-labels et IFV-Epicure. Nous avons aussi encodé sept autres échelles individuelles BBCH : céréales, légumes feuilles formant des pommes, légumes racines ou tubercules, légumes solanacées, olivier, fruits à noyaux, fruits à pépins.

4. UN CADRE MÉTHODOLOGIQUE POUR DÉCRIRE LES STADES PHÉNOLOGIQUES

Nous avons suivi la méthode *Linked Open Terms*, une méthode d'ingénierie ontologique inspirée du développement logiciel [28]. Cette méthode se concentre sur la réutilisation d'éléments (classes, propriétés objet et de type de données) existants dans des ontologies précédemment publiées sur le Web de données liées. La méthode définit des itérations sur les quatre activités suivantes : (1) spécification d'exigences ontologiques, (2) implémentation d'ontologie, (3) publication d'ontologie et (4) maintenance d'ontologie.

4.1. SPÉCIFICATION

Nos exigences ontologiques ont été spécifiées par plusieurs questions de compétence. Ces questions expriment des besoins relatifs aux échelles phénologiques de la vigne mais elles peuvent être étendues à toute échelle d'une autre culture.

- (1) Quels sont les stades d'une échelle donnée ?
- (2) Quels sont les labels français / anglais préférés et alternatifs pour un stade donné ?
- (3) Quelle est la définition en français / anglais d'un stade donné ?
- (4) Quelles plantes cultivées correspondent à l'échelle et à leurs stades ?
- (5) Quels sont les stades principaux et secondaires d'une échelle donnée de la vigne (échelle individuelle BBCH, EL, Baggiolini, IFV-labels, IFV-Epicure) ?
- (6) Quel stade principal de l'échelle individuelle BBCH est lié à un stade secondaire S de l'échelle R (EL, Baggiolini, IFV-labels, IFV-Epicure) ?
- (7) Quel stade secondaire de l'échelle individuelle BBCH est équivalent à un stade secondaire S de l'échelle R (EL, Baggiolini, IFV-labels, IFV-Epicure) ?
- (8) Quel stade secondaire de l'échelle individuelle BBCH est aligné avec un stade secondaire S de l'échelle R (EL, Baggiolini, IFV-labels, IFV-Epicure) ?
- (9) Comment sont ordonnés (suivant / précédent) les stades principaux et secondaires d'une échelle donnée ?

Pour compléter les questions de compétences 6, 7 et 8, les alignements entre les échelles ont été décrits à l'aide d'un diagramme organisant les stades par échelles phénologiques et par ordre chronologique. Ce diagramme a d'abord été enrichi par

un agronome : son but était de représenter les alignements trouvés dans la littérature. Par exemple dans la Figure 4.1 les lignes bleues claires représentent les alignements proposés par Coombe [12]. Les lignes rouges représentent les alignements proposés par Lorenz [20]. Les lignes vertes représentent les alignements proposés par Bloesh et Viret en 2008 [9]. Ensuite les deux agronomes ont proposé des alignements entre les 6 échelles. L'ontologue intervenait lors de leur discussion (1) pour clarifier le type d'alignements (exact, proche, etc.) entre deux stades, (2) pour proposer de nouveaux alignements déduits à partir des alignements précédents. Dans la figure 4.1, les traits noirs pleins représentent les alignements de type exact. Les traits noirs hachurés représentent les alignements de type proche. Le diagramme final abouti à une conceptualisation des alignements entre échelles phénologiques de la vigne.

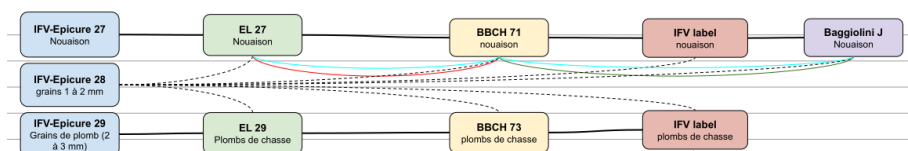


FIGURE 4.1. Extrait du schéma des alignements

4.2. IMPLÉMENTATION

Nous avons choisi de spécialiser le modèle SKOS [1] pour représenter les ensembles de stades composant une échelle phénologique ; SKOS est une recommandation du W3C. Ce modèle répond à la plupart de nos questions de compétence relatives à l'usage des labels et des alignements (Questions de compétence : 1, 2, 3, 6, 7 et 8). Un stade est représenté comme une instance de `skos:Concept`. Une échelle phénologique est représentée comme une instance de `skos:ConceptScheme`. L'ontologie sera stockée dans un fichier OWL.

4.2.1. Modèle de l'ontologie PPDO

Nous avons créé une classe `ppdo:GrowthStage`, sous-classe de `skos:Concept`. Cette classe est ensuite spécialisée en deux sous-classes, une classe pour les stades principaux et une autre pour les stades secondaires. Ces deux classes sont des classes définies pour expliciter le fait que chaque instance doit être liée à une instance de la classe `ppdo:StageDivision` par la propriété objet `ppdo:hasRank`. Ainsi toute instance de la classe `ppdo:GrowthStage` liée à l'individu `ppdo:principal`, sera automatiquement classée comme instance de la classe `ppdo:PrincipalGrowthStage` et toute instance de la classe `ppdo:GrowthStage` liée à l'individu `ppdo:secondary` sera automatiquement classée comme instance de la classe `ppdo:SecondaryGrowthStage`.

Chaque ensemble des stades phénologiques partageant le même code BBCH (par exemple le stade principal BBCH 6 ou le stade secondaire BBCH 02) sera représenté par une sous-classe de `ppdo:GrowthStage`. Une classe regroupe donc l'ensemble des

stades similaires des échelles BBCH (génériques et individuelles). Ce découpage en classe devrait permettre de retrouver facilement l'instance représentant le stade d'une échelle BBCH donnée. Ces classes répondent au besoin exprimé par la question de compétence 5. La figure 4.3 présente des exemples de classes de PPDO. Dans cet exemple, la classe intitulée `ppdo:SecondaryStage01` représente l'ensemble des stades secondaires BBCH 01 de toutes les échelles BBCH (génériques ou individuelles).

Un stade sera représenté par une instance d'une des classes de PPDO et donc une instance de `skos:Concept`. Ainsi, un stade sera décrit avec des labels multilingues préférés et alternatifs. Une échelle phénologique sera représentée comme une instance de la classe `skos:ConceptScheme`. Un stade sera lié à son échelle par la propriété `skos:inScheme`. Un stade secondaire sera lié à son stade principal par la propriété `skos:broader`.

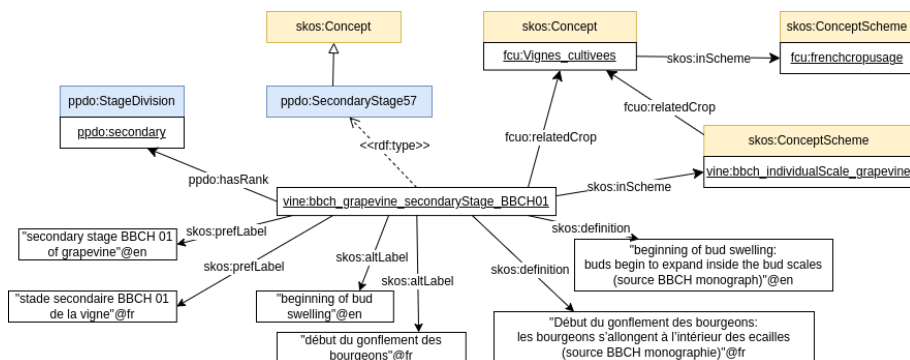


FIGURE 4.2. Représentation du stade BBCH 01 de l'échelle individuelle BBCH de la vigne

4.2.2. Modèle de l'échelle générale BBCH

Pour rappel, l'échelle générale BBCH n'est pas relative à une culture donnée. Chaque stade de l'échelle générale BBCH est représenté par une instance d'une classe spécialisant `ppdo:GrowthStage`. La figure 4.3 présente trois individus représentant des stades principaux et secondaires de l'échelle générale BBCH. Le préfixe `ppdo` signifie <https://opendata.inrae.fr/ppd-def#>. Le préfixe `ppd-gen` signifie <https://opendata.inrae.fr/ppd-res/generic#>. L'instance de la classe intitulée `ppdo:bbch_secondaryStage_BBCH01` représente le stade secondaire BBCH 01 de l'échelle générale BBCH. Cet individu est relié à l'instance de `skos:ConceptScheme` intitulée `ppd-gen:bbch_generalScale` par la propriété `skos:inScheme`.

4.2.3. Modèle de l'échelle individuelle BBCH de la vigne

Un stade de la vigne est représenté par une instance d'une des sous-classes de `ppdo:GrowthStage`. Par exemple, la Figure 4.2 présente une instance de la classe

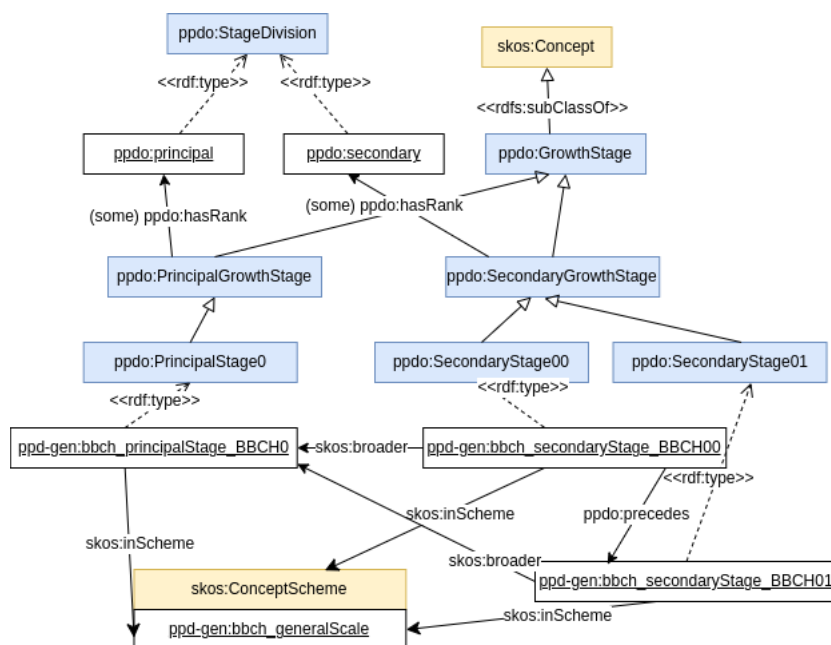


FIGURE 4.3. Extrait de l'échelle générale BBCH

ppdo:SecondaryStage01. Le préfixe vine signifie <https://opendata.inrae.fr/ppd-res/grapevine#>.

Plusieurs propriétés SKOS sont utilisées pour décrire un stade. Les labels sont déclarés à l'aide des propriétés d'annotation `skos:prefLabel` et `skos:altLabel`. Les labels préférés sont extraits de la documentation (par exemple, la monographie BBCH). Les labels alternatifs sont extraits de la définition textuelle lorsque cela est possible. Les définitions sont déclarées à l'aide de la propriété d'annotation `skos:definition`. La monographie BBCH fournit une définition par langue [21, 22, 23, 24]. La propriété objet `skos:inScheme` indique l'échelle à laquelle appartient le stade. L'échelle BBCH individuelle de la vigne est représentée par une instance de la classe `skos:ConceptScheme`.

La ou les cultures associées à un stade –ou à une échelle– sont explicitées à l'aide de la propriété objet `fcuo:relatedCrop` qui provient de l'ontologie *French Crop Usage Ontology* [2]⁽⁴⁾. Cette propriété permet de lier n'importe quelle entité (et donc un stade ou une échelle) à une culture définie dans le thésaurus *FrenchCropUsage*⁽⁵⁾. Ce thésaurus a été créé pour annoter les BSV [30]. Ce besoin correspond à la question de compétence 4.

Les liens entre les stades principaux et secondaires sont décrits à l'aide des propriétés objet `skos:broader/narrower`. La figure 4.4 illustre l'utilisation de cette propriété

⁽⁴⁾<https://agroportal.lirmm.fr/ontologies/FCUO>

⁽⁵⁾<https://agroportal.lirmm.fr/ontologies/CROPUSAGE>, (préfixe `fcu` dans les figures)

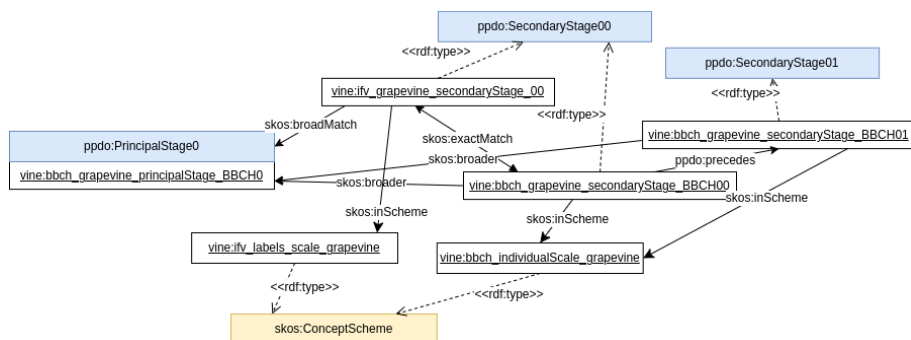


FIGURE 4.4. Alignement entre le stade BBCH 01 de la vigne et le stade correspondant dans l'échelle IFV-Labels

entre le stade principal BBCH 0 et le stade secondaire BBCH 00 de l'échelle individuelle de la vigne.

L'ordre entre les stades, tel que défini dans l'échelle, est représenté avec les propriétés objet `ppdo:precedes` et `ppdo:follows`. La figure 4.4 présente un exemple de ce lien entre le stade secondaire BBCH 00 et le stade secondaire BBCH 01 de l'échelle individuelle de la vigne. Ces propriétés permettent de répondre au besoin exprimé par la question de compétence 9.

4.2.4. Modèle des échelles produites par IFV

Nous modélisons les deux ensembles de stades produits par IFV comme deux nouvelles échelles avec la même méthodologie que celle utilisée pour représenter l'échelle individuelle BBCH de la vigne. De plus, le point d'intérêt est l'alignement des échelles entre elles. Ce besoin correspond aux questions de compétence 6, 7 et 8. Pour ce faire, nous avons utilisé les propriétés d'alignements fournies par SKOS.

Etant donné que les stades des échelles de l'IFV sont uniquement des stades secondaires, chaque stade des échelles de l'IFV est aligné à un stade principal de l'échelle individuelle BBCH de la vigne à l'aide de la propriété objet `skos:broadMatch`.

De même, chaque stade des échelles de l'IFV est aligné sur un stade secondaire de l'échelle individuelle BBCH à l'aide des propriétés `skos:exactMatch` ou `skos:closeMatch` ou `skos:broadMatch` ou `skos:narrowMatch`.

La figure 4.4 présente un exemple d'alignement entre un stade de l'échelle IFV-labels et deux stades de l'échelle individuelle BBCH. La même approche est utilisée pour aligner les autres échelles (EL, Baggiolini).

4.2.5. Encodage RDF

Une fois les modèles conçus, ils ont été encodés sous forme de classes et d'instances. Nous avons utilisé l'éditeur d'ontologie Protégé (v5.6.1) [25] avec le plugin Cellfie⁽⁶⁾. Nous avons d'abord encodé le modèle de l'ontologie définissant les classes (Figure 4.3) dans une owl:Ontology. Nous avons également ajouté des descriptions de métadonnées (par exemple, source primaire, auteurs, dates, licences) comme recommandé par [15] ou [14] en réutilisant des éléments de l'ontologie PROV du W3C [7] et du vocabulaire du Dublin Core [10].

Deuxièmement, nous avons peuplé l'ontologie avec des instances représentant tous les stades de l'échelle individuelle BBCH de la vigne, de l'échelle Baggiolini, de l'échelle EL et des deux échelles de IFV pour produire un graphe de connaissances des échelles de la vigne. Nous avons utilisé différentes sources pour peupler l'ontologie et créer le graphe de connaissances. Les monographies BBCH ont fourni les informations (labels et descriptions) en quatre langues (français, anglais, espagnol et allemand) pour l'échelle individuelle de la vigne. La publication scientifique de l'échelle de Baggiolini et de l'échelle Eichhorn–Lorenz a fourni les informations en anglais. Enfin, l'IFV nous a fourni deux fichiers CSV regroupant les informations en français et en anglais. Ensuite, nous avons extrait les données de manière semi-automatique dans différentes feuilles de calcul avant de les charger dans le graphe de connaissances de la vigne à l'aide des règles de transformation Cellfie.

Cellfie a également été utilisé pour lier et aligner les instances des classes ; les règles de transformation ont créé les liens des propriétés objet : ppdo:precedes, ppdo:follow, skos:broadener, skos:broadMatch,..., skos:exactMatch, rdf:type. Pour améliorer la cohérence de graphe de la vigne, nous avons utilisé des règles SWRL pour déduire les propriétés inverses. Une dernière vérification a été effectuée à l'aide de l'outil SKOS Play!⁽⁷⁾ : il a permis de visualiser et de contrôler le graphe SKOS et de détecter quelques erreurs.

Le fichier OWL a été utilisé pour peupler un SPARQL endpoint. Les conversions (html, ttl, jsonld, ntriple, rdf-xml) ont été effectuées à l'aide de Protégé et des outils de la librairie *raptor*⁽⁸⁾ et *python-rdflib*⁽⁹⁾. Le serveur SPARQL est jena-fuseki, de la fondation Apache⁽¹⁰⁾.

Nous parlons d'un cadre méthodologique pour décrire le résultat obtenu ; c'est-à-dire l'ontologie *BBCH based Plant Phenological Development Ontology (PPDO)* ainsi qu'un ensemble de graphes de connaissances fondées sur cette ontologie, un par culture ou ensemble de cultures proches⁽¹¹⁾. La version actuelle de l'ontologie PPDO

⁽⁶⁾Cellfie (<https://github.com/protegeproject/cellfie-plugin>) permet de transformer le contenu des feuilles de calcul en axiomes pour déclarer ou peupler une ontologie.

⁽⁷⁾<http://labs.sparna.fr/skos-play>

⁽⁸⁾<http://librdf.org/raptor/>

⁽⁹⁾<https://github.com/RDFLib/rdflib>

⁽¹⁰⁾<https://jena.apache.org/documentation/fuseki2/>

⁽¹¹⁾Produire un seul graphe de connaissances avec les instances de toutes les échelles de toutes les cultures n'aurait pas de sens pratique, les usages se faisant par culture ou sous ensemble de cultures.

est étiquetée v1.3. L'ontologie importe SKOS et PROV. Elle se compose de 571 classes, 65 propriétés objets, 7 propriétés de type de données et de 8 individus. Le tableau 4.1 présente la taille de chacun des graphes relatif à un ensemble de culture. Le graphe des 5 échelles de la vigne (ppd-gv) contient 180 individus. Le graphe des deux échelles génériques (ppd-gen) comporte 625 individus au total. Le graphe des échelles des légumes solanacées est un des plus gros avec 487 individus au total.

TABLE 4.1. La taille des graphes

prefix	nb de classes	nb de propriété objets	nb de propriété type de données	nb d'individus
ppdo	571	65	7	8
ppd-gen	571	65	7	625
ppd-gv	572	66	7	180
ppd-cr	572	66	7	89
ppd-lvfh	572	66	7	85
ppd-rsv	572	66	7	81
ppd-sol	572	66	7	487
ppd-olt	573	66	7	53
ppd-stf	572	66	7	71
ppd-pof	572	66	7	74

4.3. PUBLICATION ET MAINTENANCE DE L'ONTOLOGIE ET DES GRAPHES ASSOCIÉS

L'ontologie PPDO est accessible au public sous différentes formes comme le montre le tableau 4.2.

Tout d'abord le fichier OWL et les fichiers qui ont permis sa création sont disponibles dans un dépôt git. Le suivi des problèmes fourni par Git est l'outil utilisé pour recevoir des commentaires et contrôler les modifications. Elle est également consultable en ligne par son URI et elle est interrogeable dans un triple store. Enfin, elle est disponible sur AgroPortal [19] et sur Recherche Data Gouv par le biais de son DOI.

TABLE 4.2. Les sources de l'ontologie PPDO

prefix : ppdo
uri : <https://opendata.inrae.fr/ppd-def>
git : <https://forgemia.inra.fr/bsv/phenologicalstages>
agroportal : <http://agroportal.lirmm.fr/ontologies/PPDO>
doi : <https://doi.org/10.15454/TIMQHW>
triple store : <https://rdf.codex.cati.inrae.fr/ppd-def/sparql>

Le graphe des 5 échelles de la vigne est aussi accessible sous différentes formes voir le tableau 4.3. Pour chacun des graphes correspondant aux échelles listées dans la section 2 les mêmes sources sont définies.

TABLE 4.3. Les sources du graphe des échelles de la vigne

prefix : `ppd-gv`
uri : `https://opendata.inrae.fr/ppd-res/grapevine`
git : `https://forgemia.inra.fr/bsv/phenologicalstages`
agroportal : `https://agroportal.lirmm.fr/ontologies/PPD-GV`
doi : `https://doi.org/10.57745/YVM7VH`
triple store : `https://rdf.codex.cati.inrae.fr/ppd-res/grapevine/sparql`

5. EVOLUTION DES ÉCHELLES PAR ANNOTATION DU CORPUS

Le vocabulaire est constitué par des experts, il reflète leur vision sur les stades importants du développement de la vigne. De la même manière, les labels posés sur ces stades présentent une certaine façon d'exprimer chaque stade de développement. La constitution d'un corpus et l'analyse des expressions faisant référence aux stades phénologiques à l'intérieur de ce corpus permettent de se rendre compte des expressions qui correspondent effectivement à l'usage (à la fois en terme des stades qui sont considérés comme suffisamment significatifs pour être mentionnés, et concernant la manière dont ces stades sont exprimés). L'exploration du corpus est rendue possible par une annotation qui vise à identifier les mentions, c'est-à-dire les portions de texte, qui réfèrent à des stades phénologiques. Cette annotation est réalisée suivant deux méthodes :

- une annotation automatique, effectuée par des outils informatiques ;
- une annotation manuelle, réalisée par des experts de la vigne.

Ces deux types d'annotations répondent à des motivations différentes. L'annotation automatique permet de traiter facilement l'ensemble du corpus, assurant ainsi une large couverture. L'exploitation des résultats permet d'évaluer le vocabulaire en réalisant des premières analyses statistiques sur les stades qui ressortent le plus souvent, et les labels qui sont privilégiés à l'usage. L'annotation manuelle quant à elle répond également à d'autres besoins. Premièrement, elle permet de vérifier la complétude du lexique en mettant en évidence les expressions qui n'ont pas fait l'objet d'une description par un label dans le vocabulaire. Elle vient aussi compléter l'annotation automatique en fournissant un jeu de données d'évaluation, donnant un indicateur de qualité de la chaîne de traitement automatique, et des types d'erreurs ou d'omissions opérées. Enfin, elle propose un jeu de données d'entraînement utile pour mettre en place une chaîne de traitement supervisée, basée sur des exemples d'apprentissage.

Nous commençons par présenter le corpus sur lequel ces annotations sont effectuées en 5.1, puis décrivons les deux méthodes d'annotation, automatique et manuelle, dans la section 5.2, avant de présenter quelques résultats dans la section 5.3.

5.1. MATÉRIELS

Le corpus des BSV sélectionné porte sur la région Alsace et couvre la période 2011-2022. Le corpus complet comporte 187 bulletins. A l'intérieur de ce corpus,

un sous-corpus de 15 bulletins a été sélectionné pour l’annotation manuelle. Pour ce sous-corpus, les bulletins ont été échantillonnés aléatoirement en les distribuant équitablement à l’intérieur de 5 intervalles temporels⁽¹²⁾, afin d’assurer que chaque période de l’année soit couverte. Le corpus contient deux types de bulletins : les bulletins hebdomadaires et des bulletins bilans, publiés en fin de campagne, qui offrent une synthèse des événements ayant eu lieu sur l’année.

5.2. MÉTHODE

La reconnaissance des stades ainsi que leur normalisation sont réalisées au moyen d’une chaîne de traitement AlvisNLP [4]. Cette chaîne de traitement applique principalement les lexiques sur le texte, et est enrichie de quelques règles permettant de détecter des variations dans la façon dont sont exprimés les stades. Le code correspondant est disponible dans le dépôt Git associé <https://forgemia.inra.fr/bsv/corpus-bsv/-/tree/roia-2024>.

L’annotation manuelle quant à elle est réalisée par un groupe d’annotateurs experts en viticulture (chercheurs en agronomie et techniciens de recherche en agrosystèmes viticoles). Chaque annotateur a pour tâche de vérifier l’annotation automatique et de la corriger lorsque c’est nécessaire. Cela a nécessité un temps de formation, pour familiariser les annotateurs avec les consignes d’annotations⁽¹³⁾, s’assurer de leur compréhension, ainsi que pour leur permettre de prendre en main l’interface d’annotation AlvisAE [27]. Lorsque les consignes d’annotations ne suffisaient pas en elles mêmes à répondre aux interrogations des annotateurs (par exemple lors de la normalisation d’un stade dont le label n’était pas répertorié dans les lexiques), ceux-ci ont pu poser directement leurs questions aux personnes en charge de la campagne d’annotation, afin d’obtenir un consensus.

Au total nous estimons que l’annotation dans sa globalité a demandé 8 jours de travail. Ce temps de travail ne tient pas seulement compte de l’annotation des stades phénologiques mais également de l’annotation d’autres types d’entités d’intérêt (Maladies, Ravageurs, Organismes Nuisibles, Lieux, Dates...) ainsi que de l’annotation de relations sémantiques entre ces entités⁽¹⁴⁾. Suite à cette annotation manuelle, une phase de correction semi-automatique incluant des vérifications classiques, a fait ressortir un certain nombre d’erreurs, omissions et incohérences qui ont ensuite été manuellement corrigées.

⁽¹²⁾Une exception a été faite pour certains documents, dont la conversion du pdf au html avait produit des résultats inexploitable, et qui ont été ré-échantillonnés.

⁽¹³⁾Le guide d’annotation, qui présente le schéma et répertorie les consignes d’annotations est disponible à l’adresse <https://doi.org/10.57745/I5YVJH>.

⁽¹⁴⁾L’annotation de ces relations est particulièrement coûteuse, étant donné que ce sont des relations n-aires, mettant en jeu plusieurs entités, et qu’elles nécessitent de faire appel à un contexte large au sein du document.

5.3. RÉSULTATS

Comme indiqué précédemment, l'annotation automatique nous permet de nous faire une première idée de la fréquence d'apparition des labels, et des stades qui leur sont associés, dans les bulletins. La figure 5.1 présente les formes de surface associées à leur nombre d'occurrence respectif. Ainsi les formes « fermeture de la grappe », « floraison » et « véraison » sont les plus fréquentes avec respectivement 107, 84 et 75 occurrences, représentant à elles seules près de 30 % des 900 mentions de stades. On peut également observer les fréquences d'occurrence des stades, en regroupant cette fois-ci les mentions qui correspondent à un ensemble de stades alignés, comme dans la figure 5.2. Ici, la normalisation nous permet de neutraliser les variations dues à la forme de surface. Par exemple, les mentions « grains de pois » et « petits pois » qui désignent un même stade (IFV Epicure 31 et les stades équivalents dans les autres échelles) sont comptabilisés ensemble.

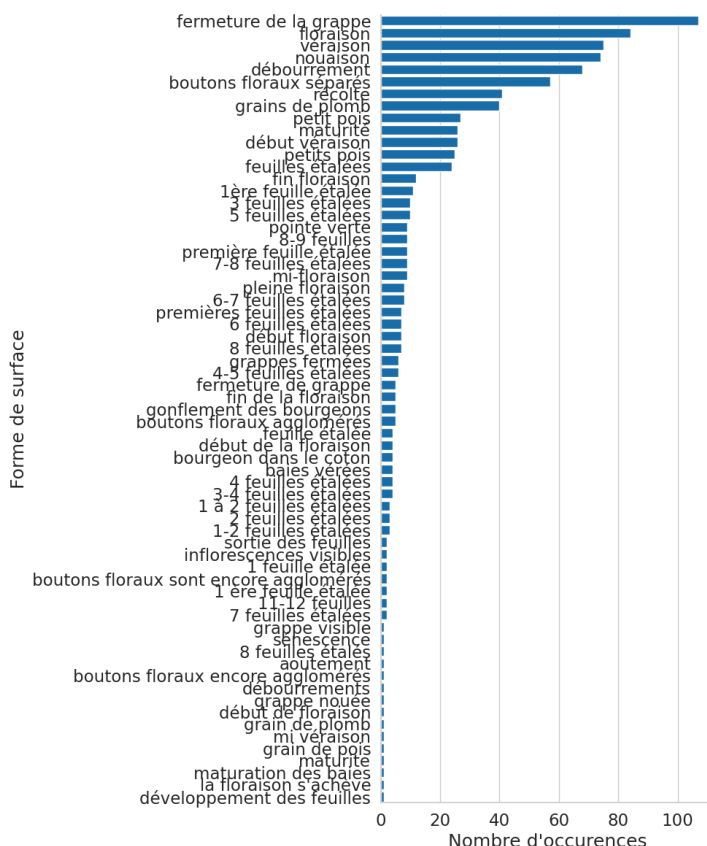


FIGURE 5.1. Fréquence des formes de surface de stades phénologiques dans le corpus annoté automatiquement.

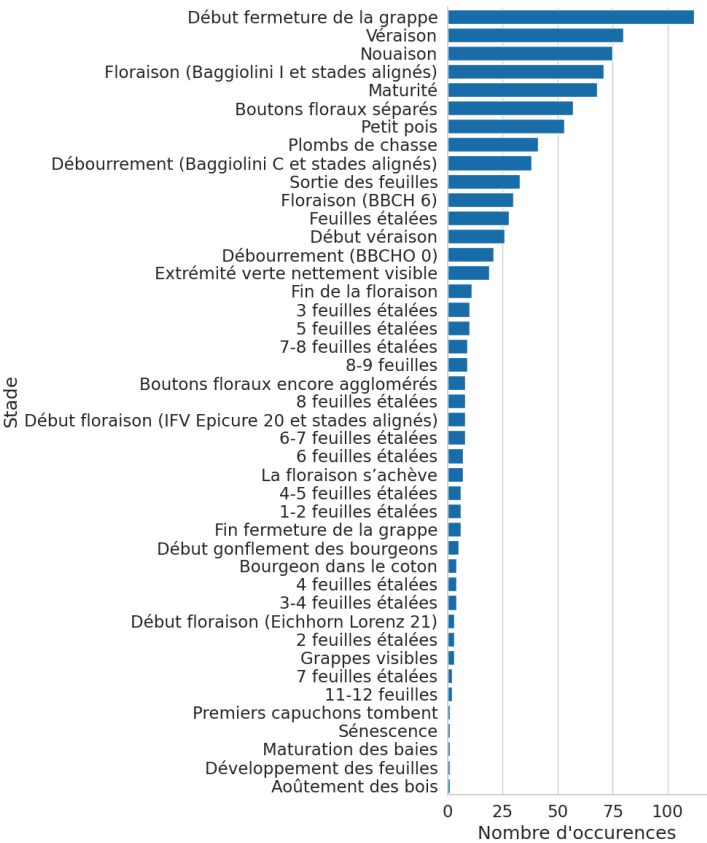


FIGURE 5.2. Fréquence des stades phénologiques normalisés dans le corpus annoté automatiquement.

Parmi les stades les plus fréquemment utilisés dans les BSV, on retrouve des stades pertinents pour la protection de la vigne face aux bioagresseurs. En particulier, les stades qui marquent le début et la fin des périodes de sensibilité à certaines maladies comme le mildiou et l'oïdium, ressortent considérablement, ce que peut confirmer l'analyse du contexte d'apparition de ces mentions.

L'annotation manuelle nous a également permis de noter un certain nombre d'expressions nouvelles qui n'apparaissaient pas dans les lexiques. Au total, on compte une soixantaine d'expressions non répertoriées. Parmi celles-ci, la très grande majorité présentent seulement une légère variation par rapport à un label existant. On retrouve par exemple un certain nombre de raccourcis, notamment « éclatement » correspondant au label connu « éclatement des bourgeons » ou « coton » pour « bourgeon dans le coton », voire des abréviations comme « BFS » pour « boutons floraux séparés ». Quelques expressions se distinguent cependant. Ainsi, certaines expressions viennent situer le stade actuel par rapport à un stade clé, *e.g.* « préfloraison » dont l'interprétation

peut varier selon le contexte mais qui correspond le plus souvent au stade « boutons floraux séparés », ou encore « préfermeture » qui correspond au stade IFV Epicure 32 « grains 5 à 6 mm ». Enfin, d'autres expressions ne s'apparentent à aucun des labels décrits. C'est l'exemple du terme « assouplissement des pellicules », typique des régions viticoles comme l'Alsace dominées par les variétés de raisins blancs, qui fait référence à un stade au tout début de la véraison, durant lequel les pellicules des baies de raisin s'assouplissent (un terme relatif à la dureté de celles-ci qui deviennent moins rigides et plus élastiques). Dans les échelles classiques, les différents stades qui rendent compte de la véraison se concentrent sur la description de la couleur des baies de variétés noires qui change au cours de la véraison, mais ne prennent pas en compte cet aspect de dureté de la pellicule.

6. DISCUSSION ET CONCLUSION

Cette transformation (RDFisation) des échelles phénologiques de la vigne est un travail collaboratif entre des experts de l'IFV et des ontologues. Le résultat améliorera la cohérence, l'usage et l'interopérabilité des échelles phénologiques actuellement existantes. Par exemple, les labels des échelles de l'IFV ont été améliorés. En effet, les labels français et les définitions ont été traduits en anglais grâce à la monographie BBCH. De plus, certaines incohérences et erreurs ont été détectées. Par exemple, l'échelle IFV-Labels mentionnait une référence au stade BBCH 88 de la vigne qui n'existe pas dans la monographie BBCH. Au cours de la phase de spécification des alignements les experts de l'IFV ont décidé d'enrichir l'échelle IFV-Epicure en ajoutant de nouveaux stades. Les alignements entre les stades des différentes échelles ont été clarifiés par l'emploi des propriétés SKOS. Les 5 échelles ont été alignées entre elles alors qu'il n'existait jusqu'à présent que des alignements partiels entre les échelles prises deux à deux.

Nous avons peuplé l'ontologie PPDO avec toutes les échelles de la vigne préconisées par l'IFV et sept échelles individuelles BBCH relatives aux céréales, aux légumes et fruits. À l'avenir, nous prévoyons de publier et d'aligner des échelles autres que BBCH relatives aux céréales (par exemple, Zadoks [36]) en utilisant le même cadre méthodologique.

La combinaison de SKOS et d'OWL permet de regrouper les instances similaires dans une classe pour faciliter leur recherche ultérieure. L'ensemble des stades d'une échelle phénologique donnée est représenté dans un modèle SKOS. En effet, la notion de Scheme au sens de SKOS permet d'identifier une échelle phénologique et de lui associer des métadonnées spécifiques (auteurs, dates, documents sources, etc.). Le mélange des hiérarchies owl:subClassOf et skos:broader/skos:narrower permet de représenter respectivement des groupes de stades principaux et secondaires et des liens hiérarchiques entre les stades d'une même échelle.

Lors du projet D2KAB, nous avons publié sur le Web de données liées notre archive BSV et les annotations associées dont les annotations de stades phénologiques [35]. De plus, les URI des stades sont utilisés dans la description de variables scientifiques étudiées dans les expérimentations de phénotypage menées à l'INRAE [26]. Ensuite,

afin d'améliorer la découverte et l'interopérabilité de l'ontologie PPDO, nous alignerons formellement les classes de PPDO avec des concepts du thésaurus Agrovoc représentant ces ensembles de stades (pour le moment nous avons déclaré simplement des liens `rdf:seeAlso`). Les alignements pourraient être complétés en alignant PPDO avec PPO. Le type d'alignement devra être étudié avec soin car les points de vue de ces deux ontologies divergent : PPO et PO représentent des processus de développement précis de chaque organe végétal. En revanche, les stades des échelles phénologiques ne sont qu'un instantané de certaines phases typiques de développement des plantes. Ces stades sont utiles pour planifier les pratiques agricoles.

Dans le cadre de l'évolution vers l'agriculture numérique, une autre perspective pour ce travail serait de rendre plus interopérables les données produites. Par exemple, en viticulture, par l'arrivée prochaine sur le marché de piquets connectés capables d'observations fines et répétées transmises en temps réel. Des prises de vue quotidiennes permettront prochainement de caractériser très finement la croissance et le développement des plantes voisines de ces postes d'observations par des algorithmes de reconnaissances de forme à partir d'images appropriées. De plus, la date de détection des stades par culture pourrait être utilisée par d'autres logiciels de traçabilité gérant par exemple le calendrier des traitements.

7. REMERCIEMENTS

Les auteurs remercient pour leur travaux d'annotation des Bulletins de Santé du Végétal Vigne : Nathalie Smits et Yvan Bouisson de l'UMR SYSTEM du centre INRAE Occitanie Montpellier. Sont aussi remerciés Marc Raynal, Marc Vergne et Christian Debord de l'Institut Français de la Vigne et du Vin du vinopôle Bordeaux Aquitaine, pour leur expertise dans l'usage et l'alignement des échelles phénologiques de la vigne. Sont remercié également plusieurs experts agricoles ou agronomes pour leur aide dans la validation des graphes par cultures : Florence Amardeilh d'Elzeard pour les légumes, Matthieu Hirschy de Acta pour les céréales, Anne Tireau de MISTEA INRAE pour l'arboriculture et Éric Garnier du CEFE pour les oliviers. Les auteurs remercient la plateforme Migale pour avoir fourni les ressources nécessaires à l'exécution des services AlvisNLP (MIGALE, INRAE, 2020. Migale bioinformatics Facility, doi : 10.15454/1.5572390655343293E12).

BIBLIOGRAPHIE

- [1] M. ALISTAIR & B. SEAN, « SKOS Simple Knowledge Organization System », W3C Recommendation, W3C, 2009.
- [2] F. AMARDEILH, S. AUBIN, S. BERNARD, C. FARON, S. BRAVO, R. BOSSY, F. MICHEL, J. RAPHEL & C. ROUSSEY, « Combining different points of view on plant descriptions: mapping agricultural plant roles and biological taxa », *Frontiers in Artificial Intelligence* **6** (2023), article no. 118803.
- [3] R. ARP, B. SMITH & A. D. SPEAR, *Building ontologies with basic formal ontology*, MIT Press, 2015.
- [4] M. BA & R. BOSSY, « Interoperability of corpus processing workflow engines : the case of AlvisNLP/ML in OpenMinTeD », in *Meeting of working Group Medicago sativa*, 2016.
- [5] M. BAGGIOLINI, « Les stades repères dans le développement annuel de la vigne et leur utilisation pratique », *Revue Romande d'Agriculture, de Viticulture et d'Arboriculture* **1** (1952), p. 4-6.

- [6] M. BAILLOD & M. BAGGIOLINI, « Les stades repères de la vigne », *Rev. Suisse Vitic. Arboric. Hortic* **25** (1993), n° 1, p. 7-9.
- [7] K. BELHAJAME, J. CHENEY, D. CORSAR, D. GARIJO, S. SOILAND-REYES, S. ZEDNIK & J. ZHAO, « PROV-O : The PROV Ontology », Tech. report, W3C, 2012.
- [8] C. BIZER, T. HEATH & T. BERNERS-LEE, « Linked Data – The Story So Far », *Semantic Web and Information Systems* **5** (2009), n° 3, p. 1-22.
- [9] B. BLOESCH, O. VIRET et al., « Stades phénologiques repères de la vigne », *Rev Suisse Vitic Arboric Hortic* **40** (2008), n° 6, p. 1-4.
- [10] D. U. BOARD, « DCMI Metadata Terms Recommendation », Tech. report, Dublin Core Metadata Initiative, 2020.
- [11] P. CLASTRE, C. PICHOT & A. CHANZY, « Vocabulary for analysis and experimentation on terrestrial and aquatic ecosystems », 2018.
- [12] B. G. COOMBE, « Growth stages of the grapevine : adoption of a system for identifying grapevine growth stages », *Australian journal of grape and wine research* **1** (1995), n° 2, p. 104-110.
- [13] L. COOPER, R. L. WALLS, J. ELSEY, M. A. GANDOLFO, D. W. STEVENSON, B. SMITH, J. PREECE, B. ATHREYA, C. J. MUNGALL, S. RENSING, M. HISS, D. LANG, R. RESKI, T. Z. BERARDINI, D. LI, E. HUALA, M. SCHAEFFER, N. MENDA, E. ARNAUD, R. SHRESTHA, Y. YAMAZAKI & P. JAISWAL, « The Plant Ontology as a Tool for Comparative Plant Anatomy and Genomic Analyses », *Plant and Cell Physiology* **54** (2012), n° 2, article no. e1.
- [14] B. DUTTA, A. TOULET, V. EMONET & C. JONQUET, « New Generation Metadata vocabulary for Ontology Description and Publication », in *11th Metadata and Semantics Research Conference, MTSR'17*, CCIS, vol. 755, Springer, 2017, p. 173-185.
- [15] D. GARIJO & M. POVEDA VILLALON, « A checklist for complete vocabulary metadata », Tech. report, WIDOCO, 2017.
- [16] H. HACK, H. BLEIHOLDER, L. BUHR, U. MEIER, U. SCHNOCK-FRICKE, E. WEBER & A. WITZENBERGER, « A uniform code for phenological growth stages of mono- and dicotyledonous plants. Extended BBCH scale, general. », *Nachrichtenblatt des Deutschen Pflanzenschutzdienstes* **44** (1992), n° 12, p. 265-270.
- [17] IFV, « Les stades phénologique de la vigne », 2017.
- [18] P. JAISWAL, L. COOPER, J. L. ELSEY, A. MEIER, M.-A. LAPORTE, C. MUNGALL, B. SMITH, E. K. JOHNSON, M. SEYMOUR, J. PREECE, X. XU, R. S. KITCHEN, B. QU, E. ZHANG, E. ARNAUD, S. CARBON, S. TODOROVIC & D. W. STEVENSON, « Planteome: A resource for Common Reference Ontologies and Applications for Plant Biology », in *24th Plant and Animal Genome Conference, PAG'16*, 2016.
- [19] A. JONQUET, CLEMENTAND TOULET, E. ARNAUD, S. AUBIN, E. DZALÉ YEUMO, V. EMONET, J. GRAYBEAL, M.-A. LAPORTE, M. A. MUSEN, V. PESCE & P. LARMANDE, « AgroPortal: a vocabulary and ontology repository for agronomy », *Computers and Electronics in Agriculture* **144** (2018), p. 126-143.
- [20] D. H. LORENZ, K. W. EICHORN, H. BLEIHOLDER, R. KLOSE, U. MEIER & E. WEBER, « Growth Stages of the Grapevine: Phenological growth stages of the grapevine (*Vitis vinifera* L. ssp. *vinifera*)—Codes and descriptions according to the extended BBCH scale », *Australian Journal of Grape and Wine Research* **1** (1995), n° 2, p. 100-103.
- [21] U. MEIER, *Entwicklungsstadien mono- und dikotyler Pflanzen: BBCH Monografie*, Open Agrar Reposeitorium, 2018.
- [22] ———, *Etapas de desarrollo de las plantas monocotiledóneas y dicotiledóneas: BBCH Monografía*, Open Agrar Reposeitorium, 2018.
- [23] ———, *Growth stages of mono- and dicotyledonous plants : BBCH Monograph*, Open Agrar Reposeitorium, 2018.
- [24] ———, *Stades phénologiques des mono-et dicotylédones cultivées : BBCH Monographie*, Open Agrar Reposeitorium, 2018.
- [25] M. A. MUSEN, « The protégé project: a look back and a look forward », *AI Matters* **1** (2015), n° 4, p. 4-12.
- [26] P. NEVEU, A. TIREAU, N. HILGERT, V. NÈGRE, J. MINEAU-CESARI, N. BRICHET, R. CHAPUIS, I. SANCHEZ, C. POMMIER, B. CHARNOMORDIC et al., « Dealing with multi-source and multi-scale information in plant phenomics : the ontology-driven Phenotyping Hybrid Information System », *New Phytologist* **221** (2019), n° 1, p. 588-601.

- [27] F. PAPAZIAN, R. BOSSY & C. NÉDELLEC, « AlvisAE: a collaborative Web text annotation editor for knowledge acquisition », in *Proceedings of the Sixth Linguistic Annotation Workshop* (N. Ide & F. Xia, eds.), Association for Computational Linguistics, 2012, p. 149–152.
- [28] M. POVEDA-VILLALÓN, « A Reuse-Based Lightweight Method for Developing Linked Data Ontologies and Vocabularies », in *9th Extended Semantic Web Conference, ESWC'12*, LNCS, vol. 7295, Springer, 2012, p. 833-837.
- [29] S. RAJBHANDARI & J. KEIZER, « The AGROVOC Concept Scheme; A Walkthrough », *Integrative Agriculture* **11** (2012), n° 5, p. 694-699.
- [30] C. ROUSSEY, S. BERNARD, F. PINET, X. REBOUD, V. CELLIER, I. SIVADON, D. SIMONNEAU & A. BOURIGAULT, « A methodology for the publication of agricultural alert bulletins as LOD », *Computers and Electronics in Agriculture* **142** (2017), p. 632-650.
- [31] C. ROUSSEY, X. DELPUECH, F. AMARDEILH, S. BERNARD & C. JONQUET, « Semantic Description of Plant Phenological Development Stages, starting with Grapevine », in *14th International Conference on Metadata and Semantics Research (MTSR2020)*, 2020.
- [32] R. SHRESTHA, E. ARNAUD, R. MAULEON, M. SENER, G. F. DAVENPORT, D. HANCOCK, N. MORRISON, R. BRUSKIEWICH & G. McLAREN, « Multifunctional crop trait ontology for breeders' data: field book, annotation, data discovery and semantic enrichment of the literature », *AoB Plants* **2010** (2010), article no. PLQ08.
- [33] B. J. STUCKY, R. GURALNICK, J. DECK, E. G. DENNY, K. BOLMGREN & R. WALLS, « The Plant Phenology Ontology : A New Informatics Resource for Large-Scale Integration of Plant Phenology Data », *Frontiers in Plant Science* (2018), article no. 517.
- [34] R. L. WALLS, L. COOPER, J. ELSE, M. A. GANDOLFO, C. J. MUNGALL, B. SMITH, D. W. STEVENSON & P. JAISWAL, « The Plant Ontology Facilitates Comparisons of Plant Development Stages Across Species », *Frontiers in Plant Science* **10** (2019), article no. 631.
- [35] N. YACOUBI AYADI, S. BERNARD, R. BOSSY, M. COURTIN, B. G. HAPPI HAPPI, P. LARMANDE, F. MICHEL, C. NÉDELLEC, C. ROUSSEY & C. FARON, « A unified approach to publish semantic annotations of agricultural documents as knowledge graphs », *Smart Agricultural Technology* **8** (2024), article no. 100484.
- [36] J. C. ZADOKS, T. T. CHANG & C. F. KONZAK, « A decimal code for the growth stages of cereals », *Weed research* **14** (1974), n° 6, p. 415-421.

ABSTRACT. — The French ANR project “Data to Knowledge in Agronomy and Biodiversity” (D2KAB) builds an archive of French agricultural alert newsletters. In order to annotate plant development stages, we need a new semantic resource. Several phenological scales already exist to describe plant development stages. The French Wine and Vine Institute (IFV) has aligned several existing scales related to grapevine. In this paper, we present our work of creating an ontological framework for semantic description of plant development stages; we introduce the *BBCH-based Plant Phenological Description Ontology*. This ontology was populated to construct several phenological scales associated with different crops: grapevine, cereals, vegetables, arboriculture. In addition, an experiment in annotating bulletins using grapevine scales was carried out to evaluate the use of those stages.⁽¹⁵⁾

KEYWORDS. — Ontology, Ontology Design, Linked data, Phenological stages, Digital Agriculture, Grapevine, BBCH, Plant Health Bulletin, Agricultural Bulletin.
