

# **Analyser l'offre alimentaire :**

## **Exploitation des listes d'ingrédients et des mentions valorisantes des produits emballés**

Ophélie Fraisier-Vannier (IRIT), Tristan Salord (IRIT),  
Marie-Benoit Magrini (INRAE, Agir), Guillaume Cabanac (IRIT, IUF)

*Séminaire AliMining, 30 septembre 2025*

# Enjeux & Défis

---

- **L'offre alimentaire répond-t-elle aux attentes de systèmes alimentaires sains et durables ?**(Rockström et al 2020, Renard et Tilman 2019)
  - Nécessité d'une plus large diversité d'espèces utilisées pour la transition agroécologique et faire face aux enjeux climatiques,
  - Nécessité de réduire la part des aliments associés à une ultra-formulation,
  - Nécessité de pratiques plus écologiques et éthiques du champ à l'industrie,
  - ...
- **Surcharge informationnelle du consommateur :**
  - Risque de perte de confiance dans une offre au final peu transparente et contrôlée,
  - Risque que les produits vertueux ne soient pas reconnus comme vrais (Magrini et al. 2025)
  - ...
- **Complexité d'analyse :**
  - Disposer de données accessibles, organisées et standardisées, comparables,
  - Pas de normes sur une large partie du descriptif des produits (intitulés des ingrédients variables ou ambigus, pas de codification universelle des procédés de transformation des ingrédients, liberté des mentions valorisantes...),
  - Un renouvellement constant dans le temps. (Ahuja et al. 2021, Sadler et al., 2021)

# Motivations

## ► Analyser massivement l'offre des produits alimentaires pour :

- Évaluer **les trajectoires de développement** de nos systèmes agroalimentaires relativement à différentes dimensions :
  - ex. Diversité des espèces utilisées vs. diversité consommée
  - ex. Types de procédés de transformation alimentaire utilisés

*« (...)identifying the species used by the food industry is of particular importance because agricultural diversity is a main lever for sustainability, and food outlets remain the main driver of cultivation choices by farmers. » Salord, Magrini et al., 2025*

- Évaluer la **qualité de l'information** :

- Identifier les produits « non-transparents », « trompeurs » vs. « sains et durables »
- ingrédients/process non identifiables.



# Enjeux méthodologiques

---

## ► Enjeux informatiques :

- Accessibilité à la donnée, où la trouver de manière systématique ?
- Comment passer de données à de l'information ?
- Comment passer d'informations à des connaissances actionnables ?
  - Création / alignement avec des vocabulaires contrôlés/ontologies
  - Coût social / environnemental des techniques de traitement de l'information

## ► Enjeux d'interdisciplinarité :

- Comment associer différentes sciences à cette analyse ?
- Peut-on construire un vocabulaire partagé ?
- Dépasser les controverses autour de l'ultra-transformation:
  - ex qu'est-ce qui qualifie un aliment ultra-transformé ?

**Le Monde**

PLANÈTE • AGRICULTURE & ALIMENTATION

**Les aliments ultratransformés ont des effets négatifs forts sur la santé en quelques semaines**

*« While the concept of ultra-processed foods has certainly entered the consumer consciousness, our results indicate that NOVA criteria do not currently allow foods to be unequivocally defined as ultra-processed. » in (Braesco, et al., 2022)*

# Où trouver la donnée ?

---



- Base de données crowd-sourcé,
- Orienté consommation
- **>= 2 millions produits**

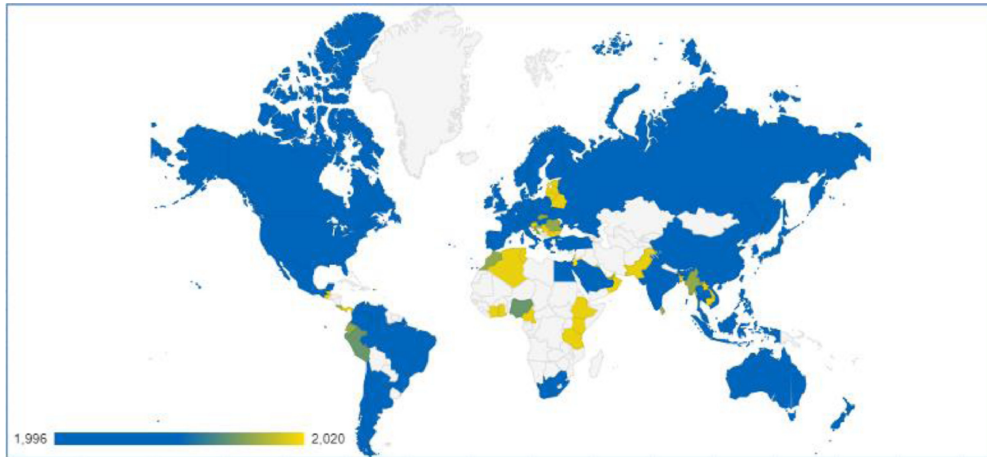


- Base de données privées
- Orienté marché et innovation
- 30000 shoppers sur le terrain
- **> 3 millions produits**



- Base de données gouvernementale,
- Orienté santé alimentaire,
- **~ 1 million produits**

# Exemple d'application sur MINTEL-GNPD



Couverture de MINTEL (86 pays)

- ▶ MINTEL :
  - ✓ couverture mondiale,
  - ✓ orientation innovation

- ▶ Focus sur une famille végétale porteuses d'**enjeux environnementaux et de santé forts** : les **légumineuses à graines (pois, haricots, lentilles, pois chiches... et soja)**
- ▶ Objectifs :
  - ✓ évaluer leur diversité (espèces) et leur progression sur les marchés,
  - ✓ caractériser leur « valeur d'usage ».

« Food diversity is a challenging issue for sustainable agrifood systems. Diets are increasingly dependent on branded packaged foods. Therefore, the crop diversity offered in the food market through these products is of particular importance. We scrutinize this diversity for some crops under great societal challenge: pulses. » - Salord, T., Magrini, M.-B., et al., 2024

# Données et méthodes

---

## ► Quelles sont les données disponibles ?

- L'**intégralité** des données du **packaging produit** dans une base de données organisée

## MAIS

- Les données ne sont **pas liées** entre elles (par exemple données de composition et liste d'ingrédients),
- Les données sont pré-traitées et standardisées = **perte d'information**,
  - Ex : liste d'ingrédients standardisées sous leur forme simple (perte de la hiérarchie, des données textuelles contextuelles, des mentions de volumes ou de poids, etc.)
- **Pas d'appariement** avec des vocabulaires contrôlés (thesaurus, ontologies, etc.) = erreurs de catégorisation de certaines espèces végétales

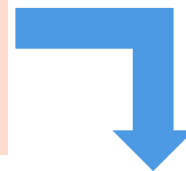
## Prendre au sérieux les « narrations produits »

- ## ► Développement d'un **parser** (« FoodCop » - « Food Composition Parser ») le moins destructif possible + **Tagger**

# Données et méthodes : FoodCOP

## Input :

tomato\* (tomato pulp\*, tomato concentrate\*), water, rehydrated red kidney beans\* (11%), courgettes\* (8e) , carrots\* (7o) , onions\* (6o) , peppers\* (6%) , soyproteins\*(5%),rehydrated chickpeas\*, cold extracted extra virgin olive oil\*, spices\*, sea salt, garlic\*, basil\*  
\*ingredients from organic farming



```
{
  '75_1': {'rawing': 'tomato', 'level': 0, 'comment': ['ingredients from organic farming']},
  '75_2': {'rawing': 'tomato pulp', 'level': 1, 'comment': ['ingredients from organic farming']},
  '75_3': {'rawing': 'tomato concentrate', 'level': 1, 'comment': ['ingredients from organic farming']},
  '75_4': {'rawing': 'water', 'level': 0},
  '75_5': {'rawing': 'rehydrated red kidney beans', 'level': 0, 'prop': ['11%'], 'comment': ['ingredients from organic farming']},
  '75_6': {'rawing': 'courgettes', 'level': 0, 'prop': ['8%'], 'comment': ['ingredients from organic farming']},
  '75_7': {'rawing': 'carrots', 'level': 0, 'prop': ['7%'], 'comment': ['ingredients from organic farming']},
  '75_8': {'rawing': 'onions', 'level': 0, 'prop': ['6%'], 'comment': ['ingredients from organic farming']},
  '75_9': {'rawing': 'peppers', 'level': 0, 'prop': ['6%'], 'comment': ['ingredients from organic farming']},
  '75_10': {'rawing': 'soy proteins', 'level': 0, 'prop': ['5%'], 'comment': ['ingredients from organic farming']},
  '75_11': {'rawing': 'rehydrated chickpeas', 'level': 0, 'comment': ['ingredients from organic farming']},
  '75_12': {'rawing': 'cold extracted extra virgin olive oil', 'level': 0, 'comment': ['ingredients from organic farming']},
  '75_13': {'rawing': 'spices', 'level': 0, 'comment': ['ingredients from organic farming']},
  '75_14': {'rawing': 'sea salt', 'level': 0},
  '75_15': {'rawing': 'garlic', 'level': 0, 'comment': ['ingredients from organic farming']},
  '75_16': {'rawing': 'basil', 'level': 0, 'comment': ['ingredients from organic farming']}
}
```

# Données et méthodes : tagger

- ▶ Ambiguïté des noms d'espèce = importance du contexte de citation
- ▶ Variabilité orthographique et culturelle des noms d'espèces = pas de vocabulaire contrôlé unique = **travail d'expertise**
- ▶ Quid des autres espèces végétales ?

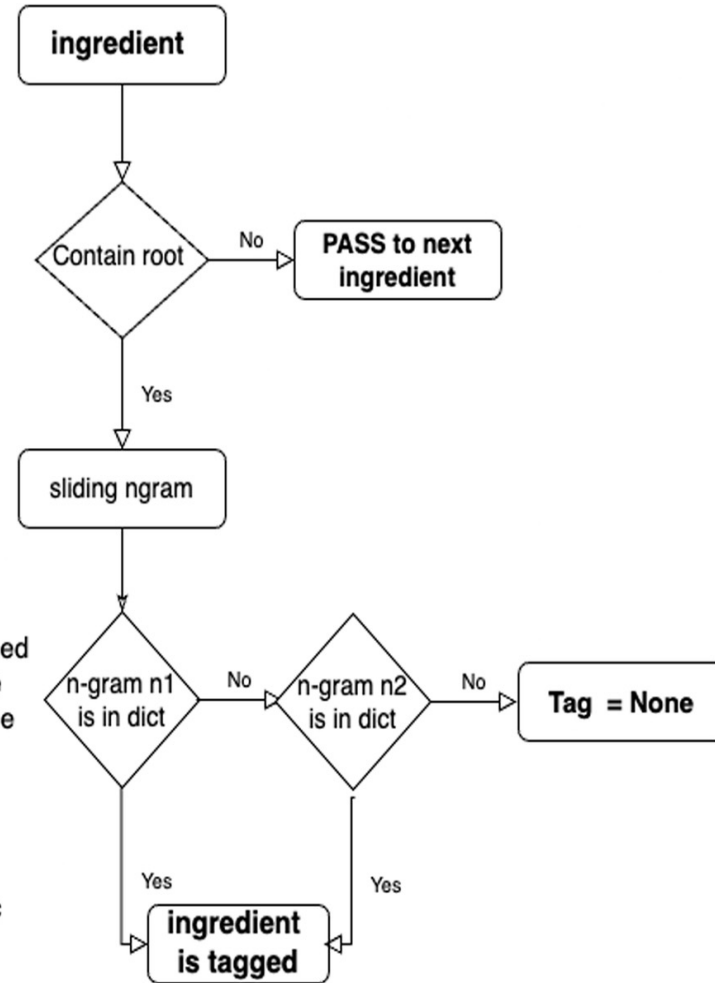
Example:  
'split yellow pea'

'split yellow pea'  
contains root 'pea'

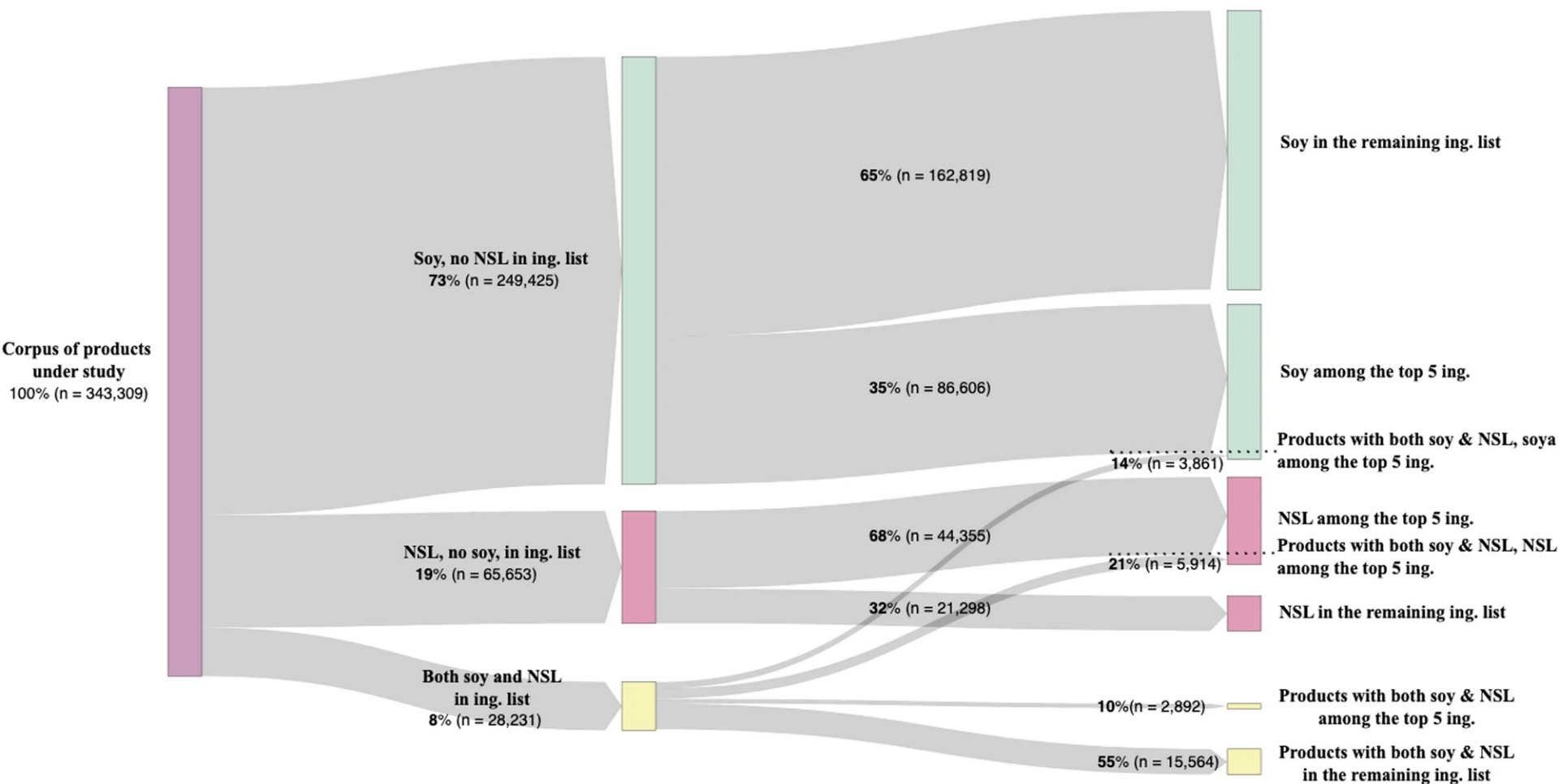
ingredient expression is splitted  
into its different parts:  
'split', 'yellow', 'pea'

For each possible n-gram contained  
in the splitted ingredient name we  
looked for a correspondance in the  
species dictionary we have built

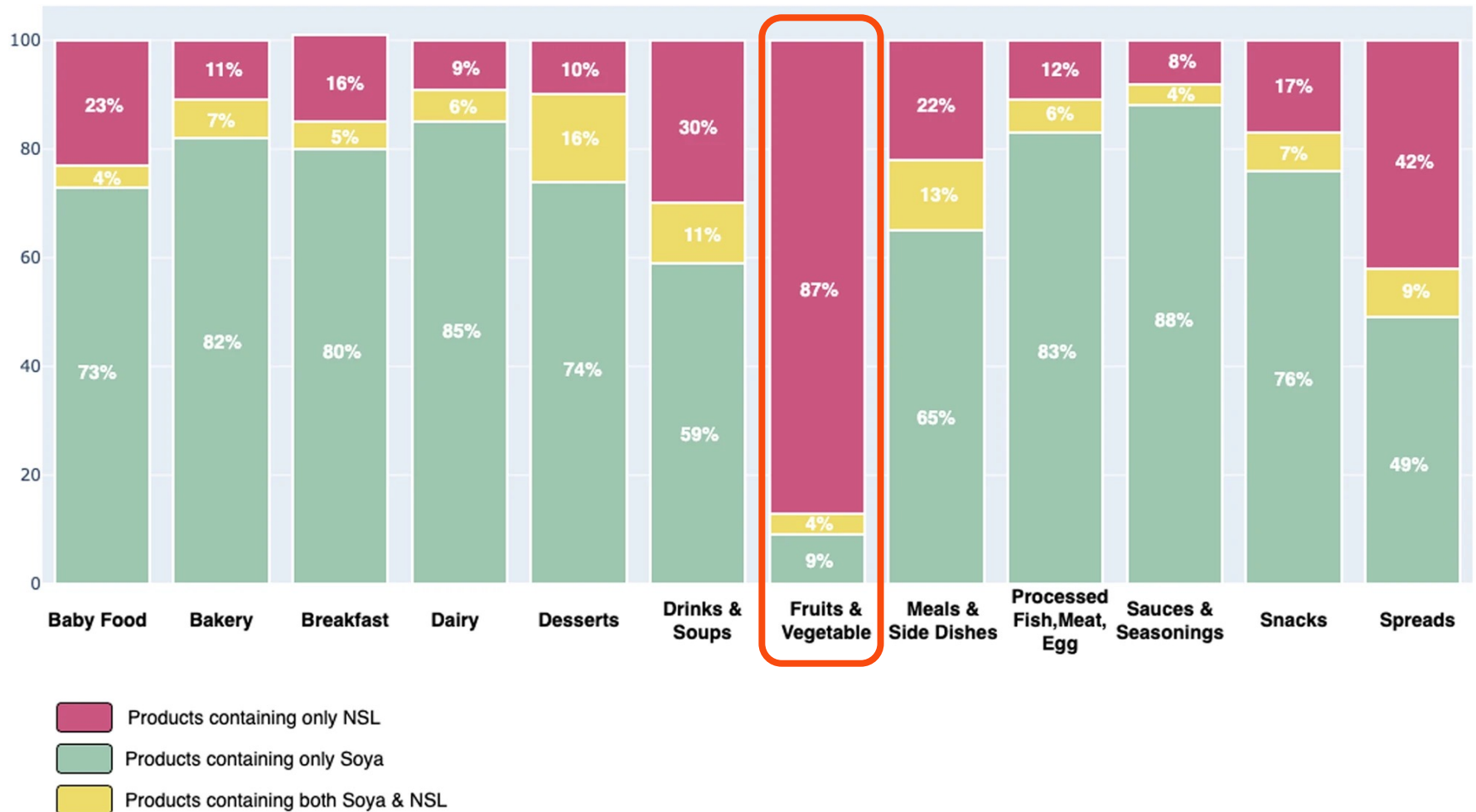
ingredient is tagged with scientific  
species name corresponding to  
the combinaison 'yellow'+ 'pea'



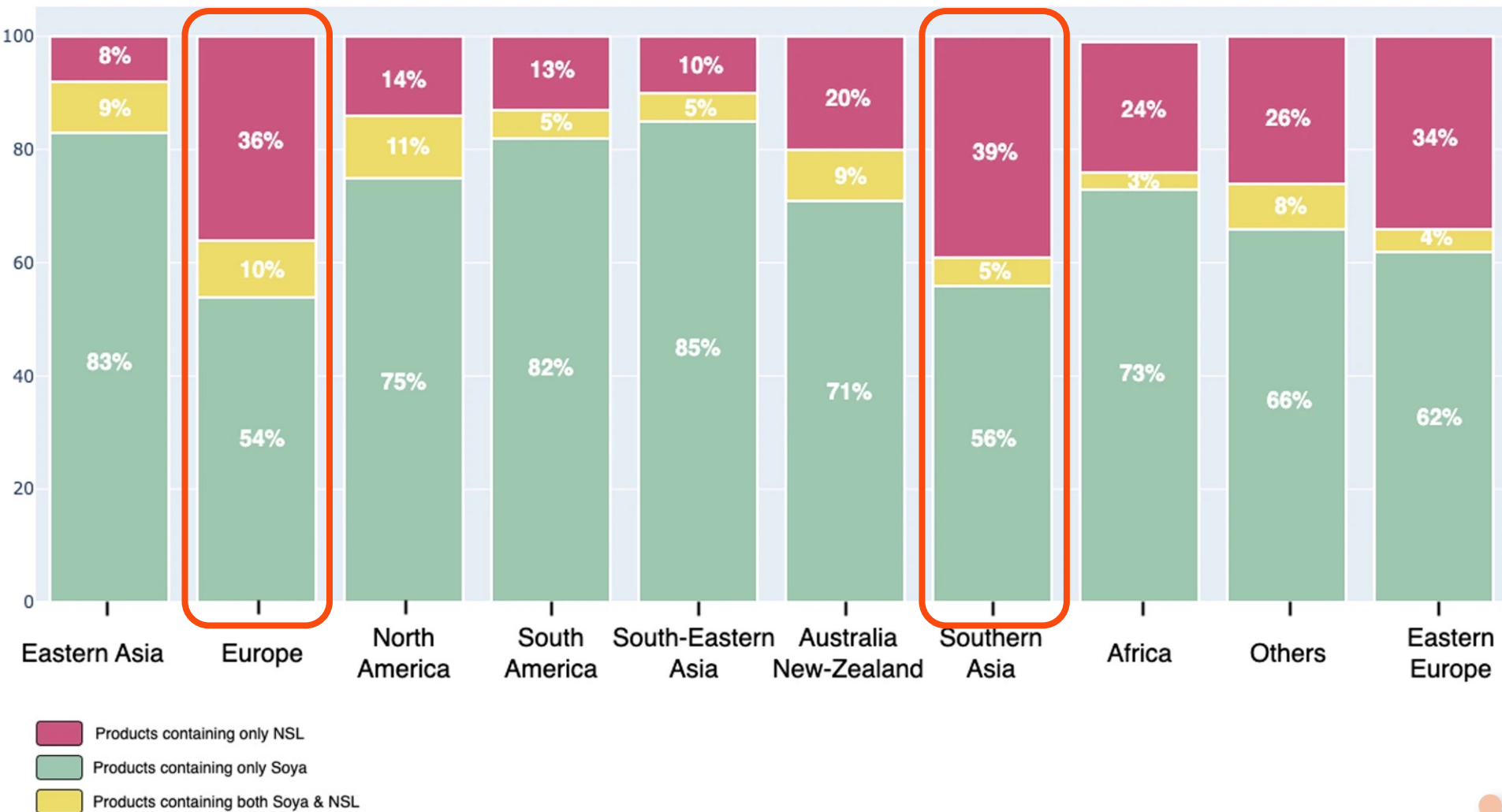
# Résultats : l'écrasante présence du soja



# Résultats : les légumineuses dans les différents segments de marché

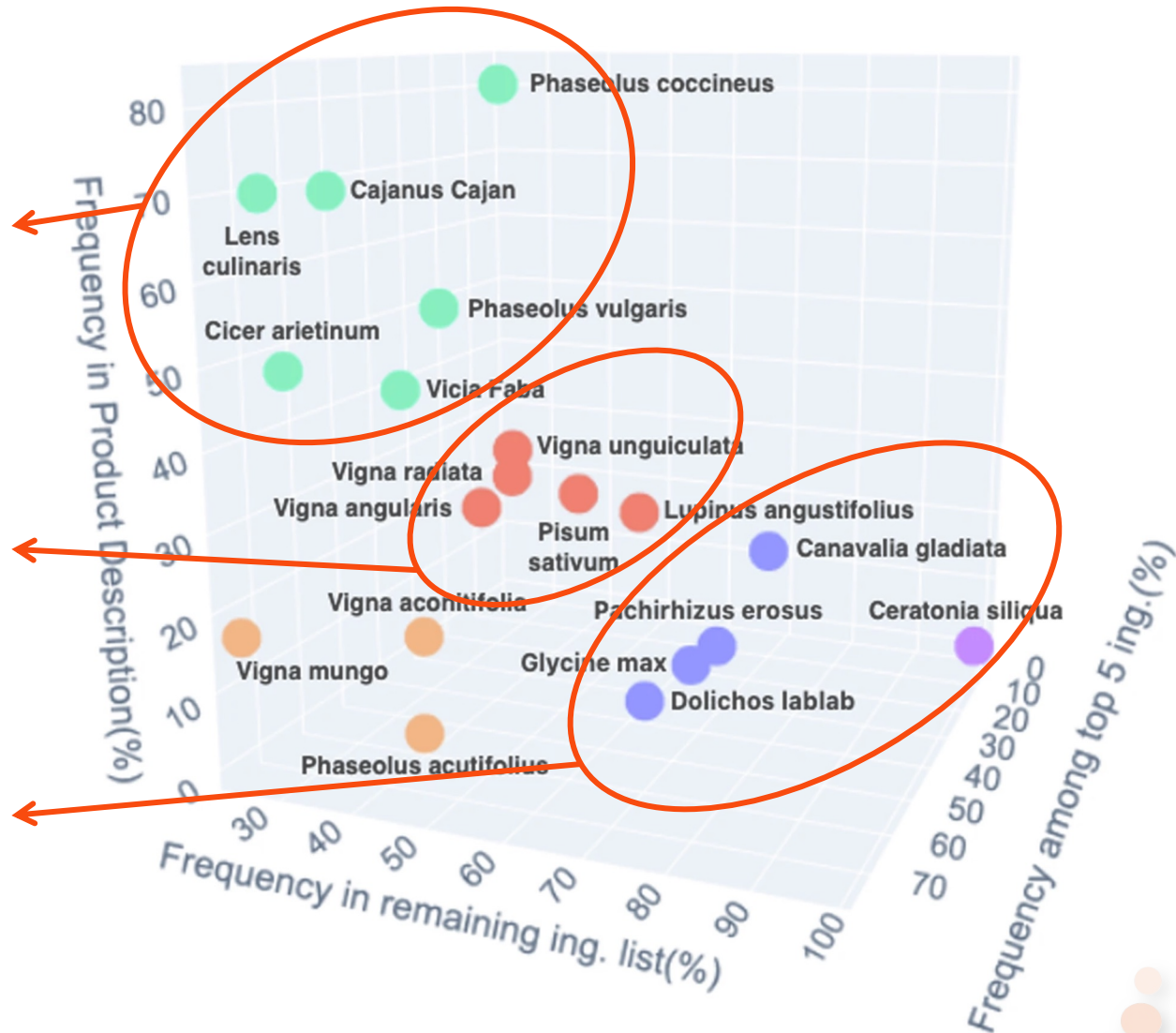


# Résultats : des ingrédients « mondialisés » ?



# Résultats : un proxy des usages des ingrédients

- Des espèces à « identité produit » forte, peu associées à des usages sous forme de fractions.
- Des espèces « discrètes » aux stratégies de valorisation très différenciées.
- Des espèces fortement associées à des usages sous forme de fractions - espèce-modèle : le soja.



# Mentions valorisantes

« Les mentions valorisantes constituent un mode de valorisation des produits agricoles et agroalimentaires pour lesquels un qualificatif spécifique est mis en exergue. »

— Rapport du 114e Congrès des notaires de France

Précision de la DGCCRF : une mention valorisante apposée sur un produit doit être  
« **fiable, non trompeuse, vérifiable, claire et justifiée par des preuves scientifiques** »



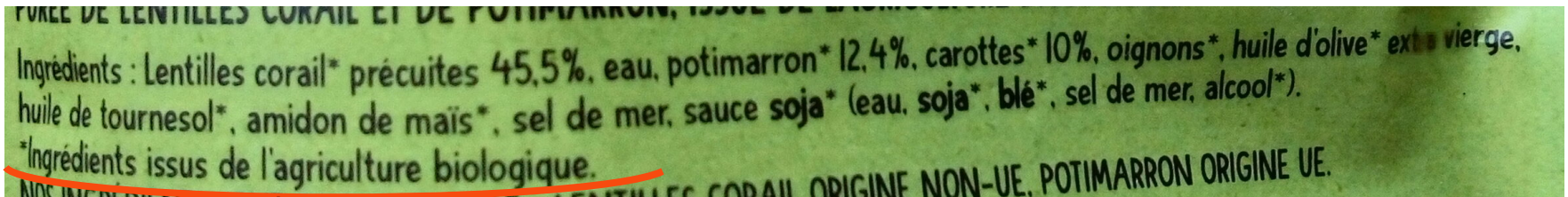
*Mentions publiques certifiées*  
→ univoques



*Mentions libres*  
→ « la jungle »  
→ ambiguïtés, termes flous  
ex : 'naturellement', 'soutient les agriculteurs'

# Mentions valorisantes

Parfois présentes dans les listes d'ingrédients



*Extraction de ces commentaires via le parser*

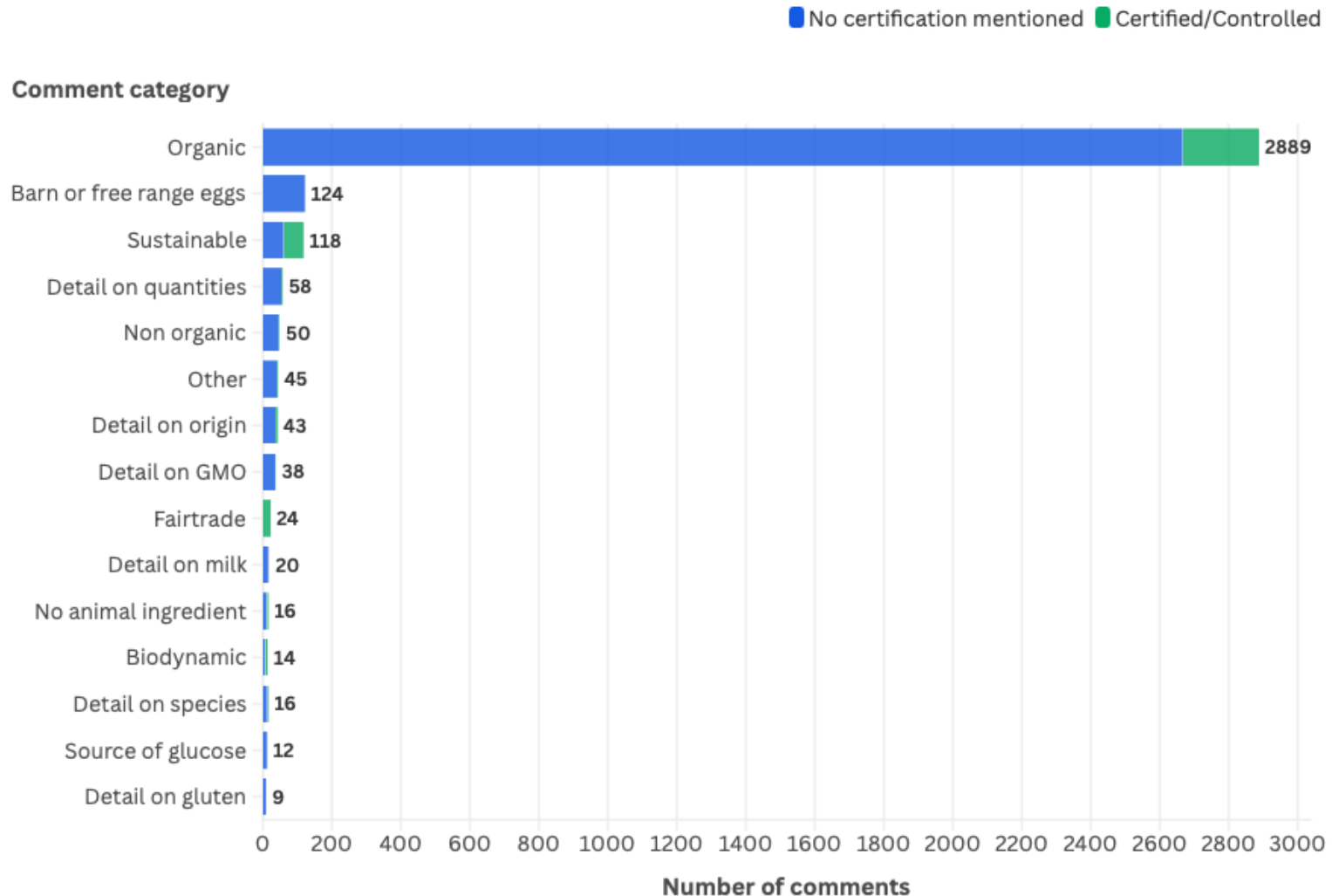
Dans notre jeu de données

- **3425 commentaires** extraits des listes d'ingrédients
- **~2800 produits** distincts



Presque exclusivement mentions valorisantes + infos additionnelles à la marge

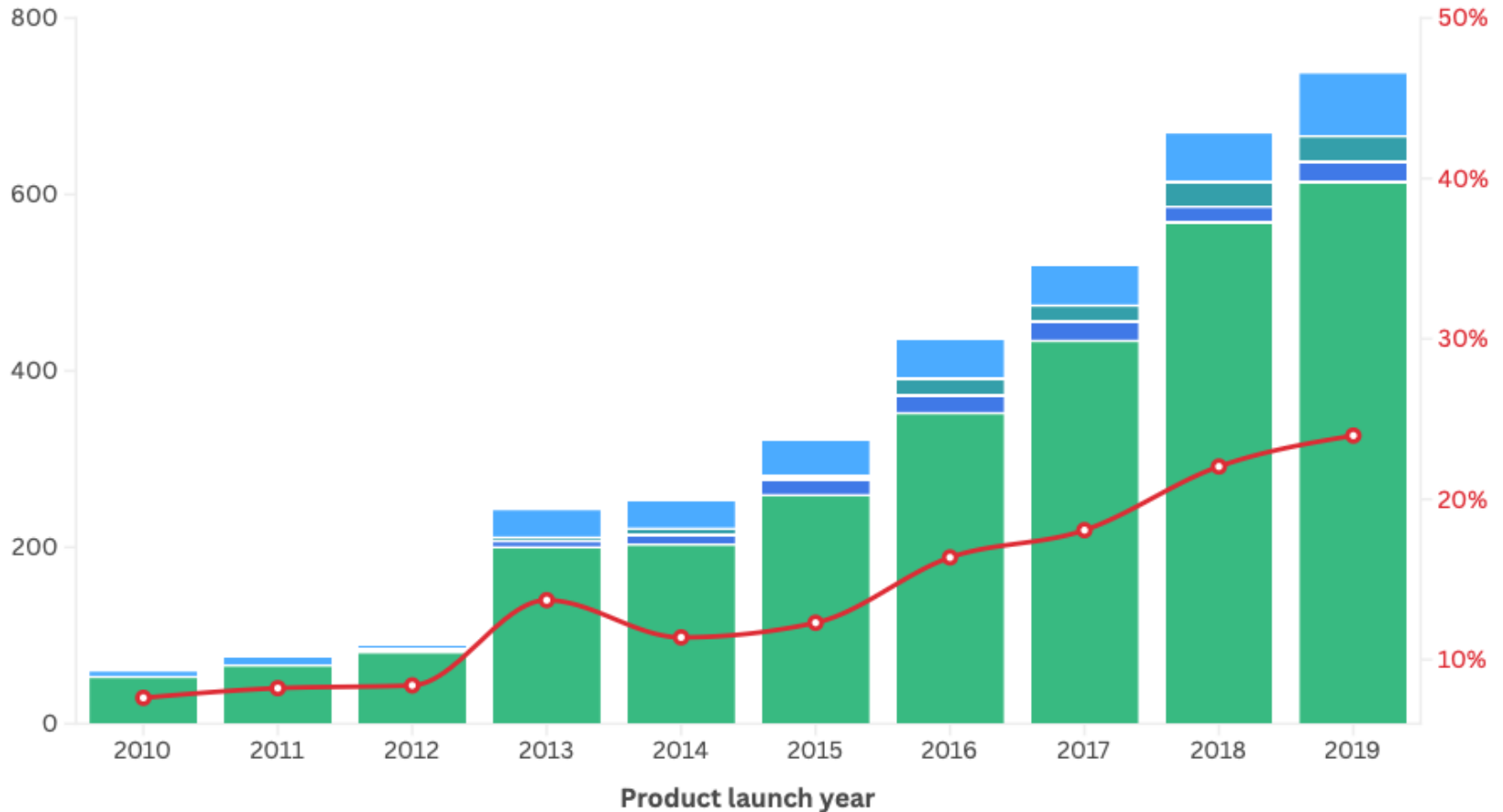
# Mentions valorisantes



# Mentions valorisantes

Organic Barn or free range eggs Sustainable Rest

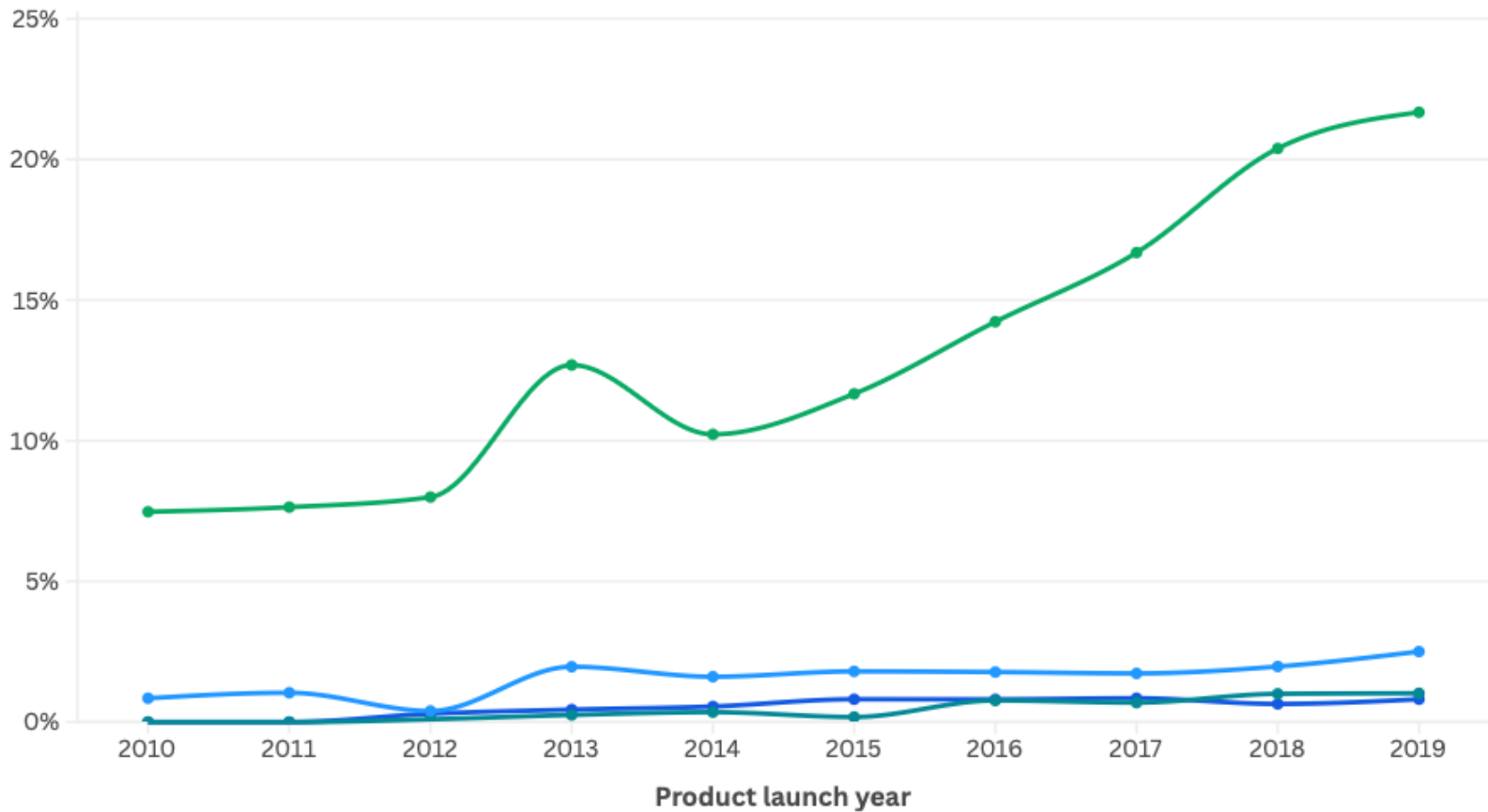
Number of comments



# Mentions valorisantes

■ Organic ■ Barn or free range eggs ■ Sustainable ■ Other comments

Proportion of product with comments



# Perspectives

---

- **Améliorations techniques / logicielles**

- Systématisation de l'appariement entre nom ingrédient et espèce végétale / animale
  - Machine learning ?
- Alignement entre listes des ingrédients & nomenclatures institutionnelles des additifs
- Exploitation des contextes de citation des ingrédients pour aider à déterminer les procédés de transformations
- Mise en corrélation *ranking ingrédients / proportion nutritionnelle*

- **Enjeux socio/éco/agro**

- Évaluation de la prévalence de certaines espèces végétales / animales / procédés de transformation sur les marchés alimentaires et des trajectoires technologiques
- Évolution des mentions valorisantes dans le temps :
  - Produits les + / les – valorisés
  - Comment est faite la mise en valeur ?
  - L'effet normatif sur les enjeux de durabilité
- Confrontation de l'offre et de la demande : nouveaux produits lancés sur le marché mais pas d'identification conduite des produits qui se vendent / qui disparaissent

# Bibliographie

---

Ahuja, J. K. C., Li, Y., Bahadur, R., Nguyen, Q., Haile, E., & Pehrsson, P. R. (2021). IngID: A framework for parsing and systematic reporting of ingredients used in commercially packaged foods. *Journal of Food Composition and Analysis*, 100, 103920. <https://doi.org/10.1016/j.jfca.2021.103920>

Braesco, V., Souchon, I., Sauvant, P., Haurogné, T., Maillot, M., Féart, C., Darmon, N., 2022. Ultra-processed foods: how functional is the NOVA system? *Eur J Clin Nutr* 76, 1245–1253. <https://doi.org/10.1038/s41430-022-01099-1>

Magrini, M.-B., Bertagnoli, S., Cadore, J.-L., Caillat, H., Carlin, F., Debaeke, P., Delaby, L., Foucher, F., Julier, B., Larzul, C., Fortun-Lamothe, L., Lebret, B., Le Gouis, J., Lullien-Pellerin, V., Nozieres-Petit, M.-O., Prache, S., Renard, M., Sebillotte, C., Schouler, C., ... Détang-Dessendre, C. (2025). The study « Agroecology & Market » carried out by the value-chain groups of INRAELe chantier « Agroécologie & Marché » conduit par les groupes filières INRAE. <https://doi.org/10.17180/CIAG-2025-VOL100-ART01>

Magrini, M.-B., Salord, T., Cabanac, G., 2023. The unbalanced development among legume species regarding sustainable and healthy agrifood systems in North-America and Europe: focus on food product innovations. *Food Sec.* 15, 187–200. <https://doi.org/10.1007/s12571-022-01294-9>

Renard, D., & Tilman, D. (2019). National food production stabilized by crop diversity. *Nature*, 571(7764), 257-260. <https://doi.org/10.1038/s41586-019-1316-y>

Rockström, J., Edenhofer, O., Gaertner, J., & DeClerck, F. (2020). Planet-proofing the global food system. *Nature Food*, 1(1), 3-5. <https://doi.org/10.1038/s43016-019-0010-4>

Sadler, C. R., Grassby, T., Hart, K., Raats, M., Sokolović, M., & Timotijevic, L. (2021). Processed food classification : Conceptualisation and challenges. *Trends in Food Science & Technology*, 112, 149-162. <https://doi.org/10.1016/j.tifs.2021.02.059>

Salord, T., Magrini, M.-B., Lullien-Pellerin, V., Cabanac, G., Amiot, M.-J., Barron, C., Boire, A., Micard, V., Weber, M., 2024. Crop diversity used in branded products with focus on legume species worldwide. *npj Sci Food* 8, 68. <https://doi.org/10.1038/s41538-024-00305-7>