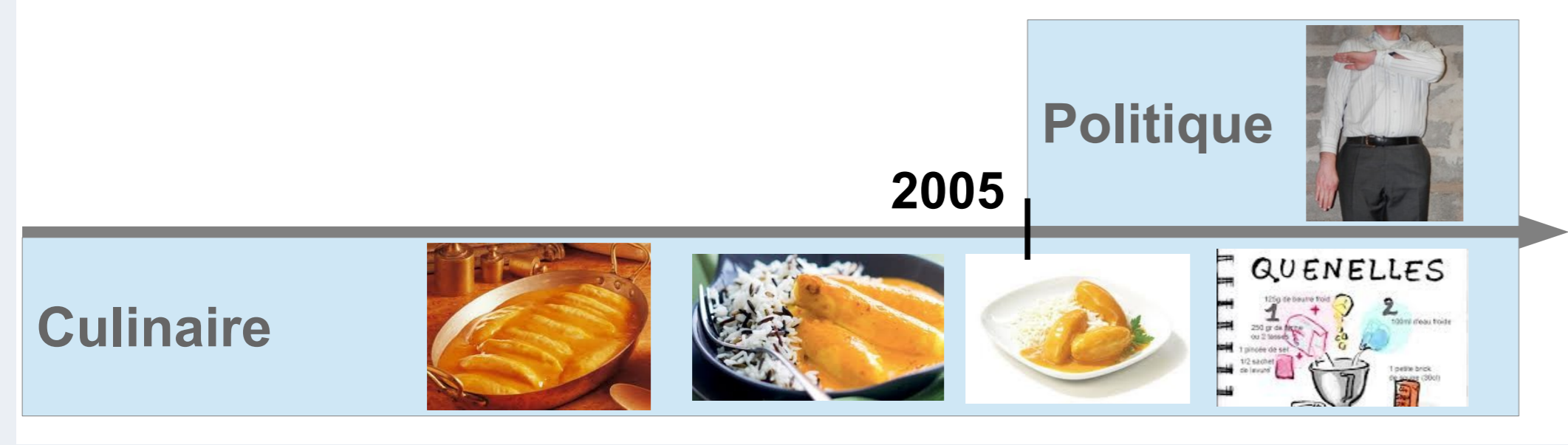


Problématique : vers la détection automatique d'un changement sémantique particulier (création métaphorique)

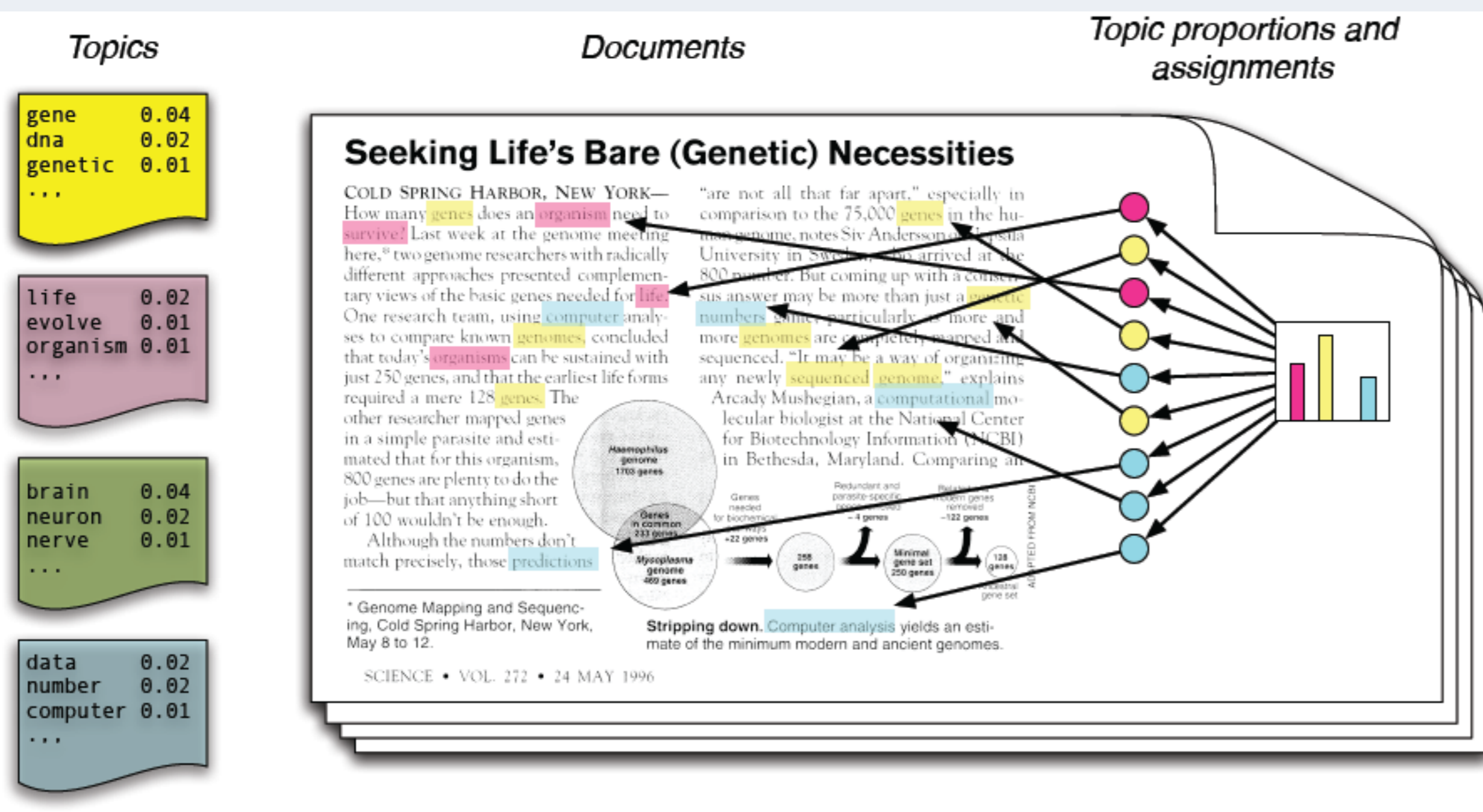
Significations du signifiant « quenelle » :



Représenter automatiquement ce changement thématique ?

- utilisation de *Topic Modeling* Lau et al. (2012, 2014)
- application aux occurrences de « quenelle » (corpus d'articles de presse)

Utilisation du Topic Modeling



Steyvers and Griffiths (2007)

Questions abordées

- alignement topics découverts ↔ significations lexicographiques de *quenelle* ?
- signification prédominante dans notre corpus ?
- indice d'émergence de la nouvelle signification ?

Réponse proposée

Modèle : HDP (Hierarchical Dirichlet Process), non-paramétré, #topics inféré
Corpus : paragraphes extraits d'articles de presses, contenant « quenelle »
Topics ≈ thèmes du paragraphe ≈ significations de « quenelle »

Ressources

Corpus

- REF 164 paragraphes du *Monde*
 - 1987-2006
 - sens *culinaire*
- NOUV 342 paragraphes de 160 articles (Le Monde, Libération, La Croix, etc.)
 - septembre 2013 - janvier 2014
 - 303 sens *geste*, 38 *culinaire*, 1 autre (manuellement)

Prétraitement : Morphalou

Dictionnaire

- Wikipedia : 2 articles
 - *quenelle* culinaire
 - *quenelle* geste

Topics appris sur corpus REF+NOUV

- T1 : *quenelle geste antisémite salut photo nazi bras humoriste français effectuer*
- T2 : *quenelle jean marier marque fion fond glisser petite sionisme déposer*
- T3 : *quenelle carte menu plat brochet cuisine rue compter veau dessert*
- T4 : *quenelle spectacle public geste jeune interdiction antisémite humoriste monde ordre*
- T5 : *quenelle brochet pain épices grand sauce chair cuisine écrevisse beurre*
- T6 : *vide quenelle couverts charges cru debout derrière docteur effet émission*

Zoom sur un paragraphe

Cette *procédure* fait suite à la diffusion d'un *reportage* de France 2 qui le *montre*, sur *scène*, s'en prenant verbalement au *journaliste* de France Inter Patrick Cohen en *évoquant* les *chambres à gaz*. L'*artiste* est déjà dans le viseur du *ministère de l'intérieur*, qui *envisage* d'*interdire* ses *spectacles*, après qu'il a *créé* la « *quenelle* », *geste* évoquant le *salut nazi*.

Distribution des topics :

T1 54.00%, T4 46.00%

Alignement $s_i \leftrightarrow$ topic t_j Lau et al. (2014)

- article Wikipédia ≈ sens dictionnaire $s_i \rightsquigarrow$ distribution de mots $S(s_i)$
- $sim(s_i, t_j)$ = divergence (Jensen-Shannon et Kulbach-Leibler) de $S(s_i)$ et distribution $T(t_j)$.

	T5	T3	T1	T2	T6	T4	moyenne
culinaire	0.266	0.2228	0.1162	0.1141	0.0714	0.0709	0.1436
geste	T1	T2	T4	T5	T3	T6	moyenne
	0.3917	0.3181	0.2559	0.1722	0.1289	0.071	0.223

Signification prédominante Lau et al. (2014)

Distribution topics t_j / significations s_i

Corpus	T1	T2	T3	T4	T5	T6
REF	0	6	118	0	36	4
NOUV	229	32	9	42	29	1
Total	229	38	127	42	65	5

$prevalence(s_i)$ = moyenne des $sim(s_i, t_j)$, pondérées par # paragraphes à t_j prédominant.

sens	prévalence	proportion paragraphes (manuellement)
culinaire	0.3625	196/506 = 0.3873
geste	0.6375	303/506 = 0.5988

Émergence de topics nouveaux Lau et al. (2012)

Lau et al. (2012)

- Score de nouveauté du topic t_i élevé si t_i
 - très fréquent dans NOUV
 - peu fréquent dans REF

	T1	T2	T3	T4	T5	T6
	108.81	1.56	-0.96	19.14	-0.61	-0.88

Bilan

Bénéfices

- méthode simple
- mesures de Lau et al. pertinentes
- création métaphorique représentée
- approche linguistique réaliste

Continuation

- appliquer à d'autres changements : *souris, caviar*...
- appliquer à d'autres types de changement sémantique ?

Problèmes

- alignement des topics avec définitions des dictionnaires ?
- détection *non-supervisée* des changements sémantiques ?

Références

- Lau, J. H., Cook, P., McCarthy, D., Gella, S., and Baldwin, T. (2014). Learning Word Sense Distributions, Detecting Unattested Senses and Identifying Novel Senses Using Topic Models. In *Proceedings of ACL 2014*, Baltimore, USA.
- Lau, J. H., Cook, P., McCarthy, D., Newman, D., and Baldwin, T. (2012). Word sense induction for novel sense detection. In *Proceedings of EACL 2012*.
- Steyvers, M. and Griffiths, T. (2007). Probabilistic topic models. *Handbook of latent semantic analysis*, 427(7).