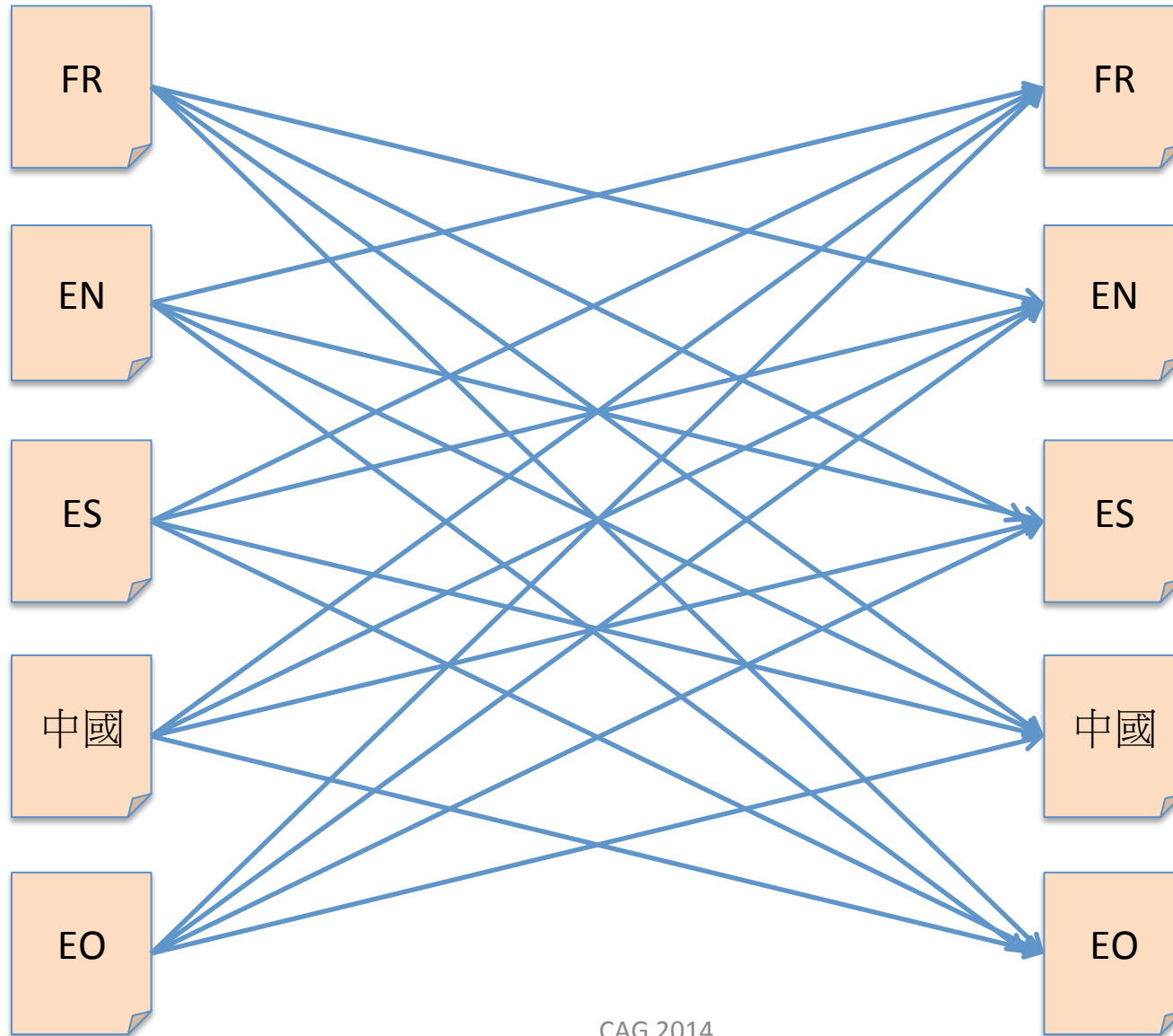


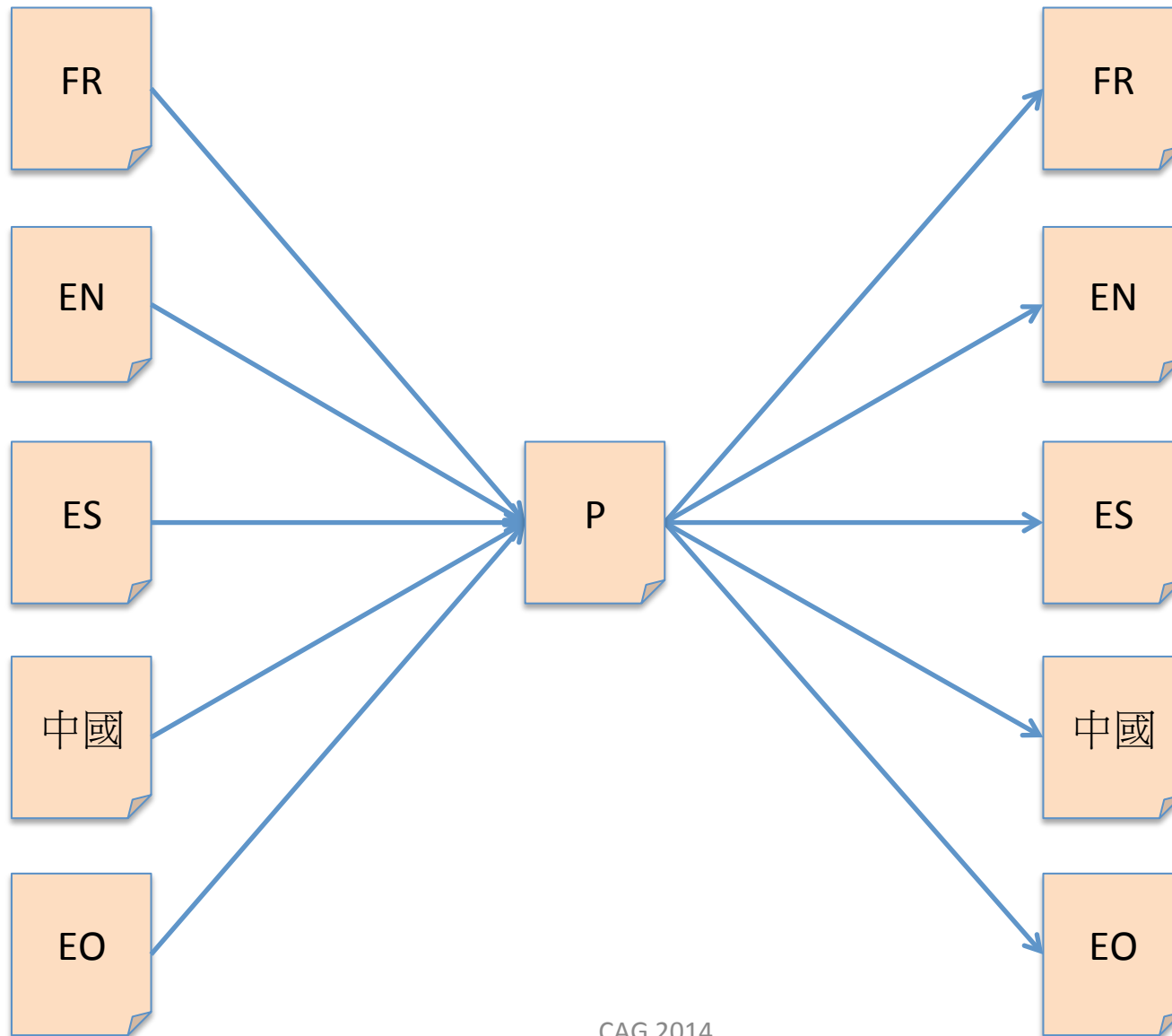
Langues Pivot

GF/CAG2014/Chancy

Problème : $N \times (N-1)$



Solution géniale : N + N



Quelle langue prendre pour P ?

N'importe quelle langue humaine

- elles sont toutes équivalentes
(avec le lexique approprié)

- mais elles sont toutes ambiguës
=> perte de sens

Perte d'information

EN

He saw the girl with **his** telescope

FR

Il a vu la fille avec son télescope

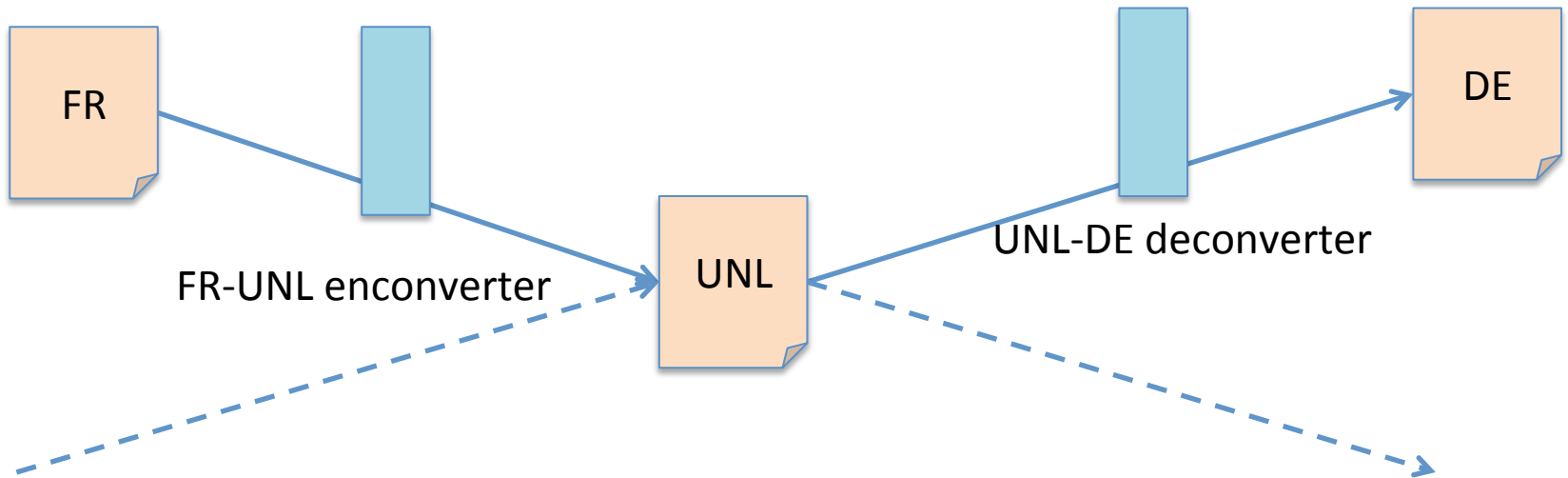
Le *son* français est ambigu

DE

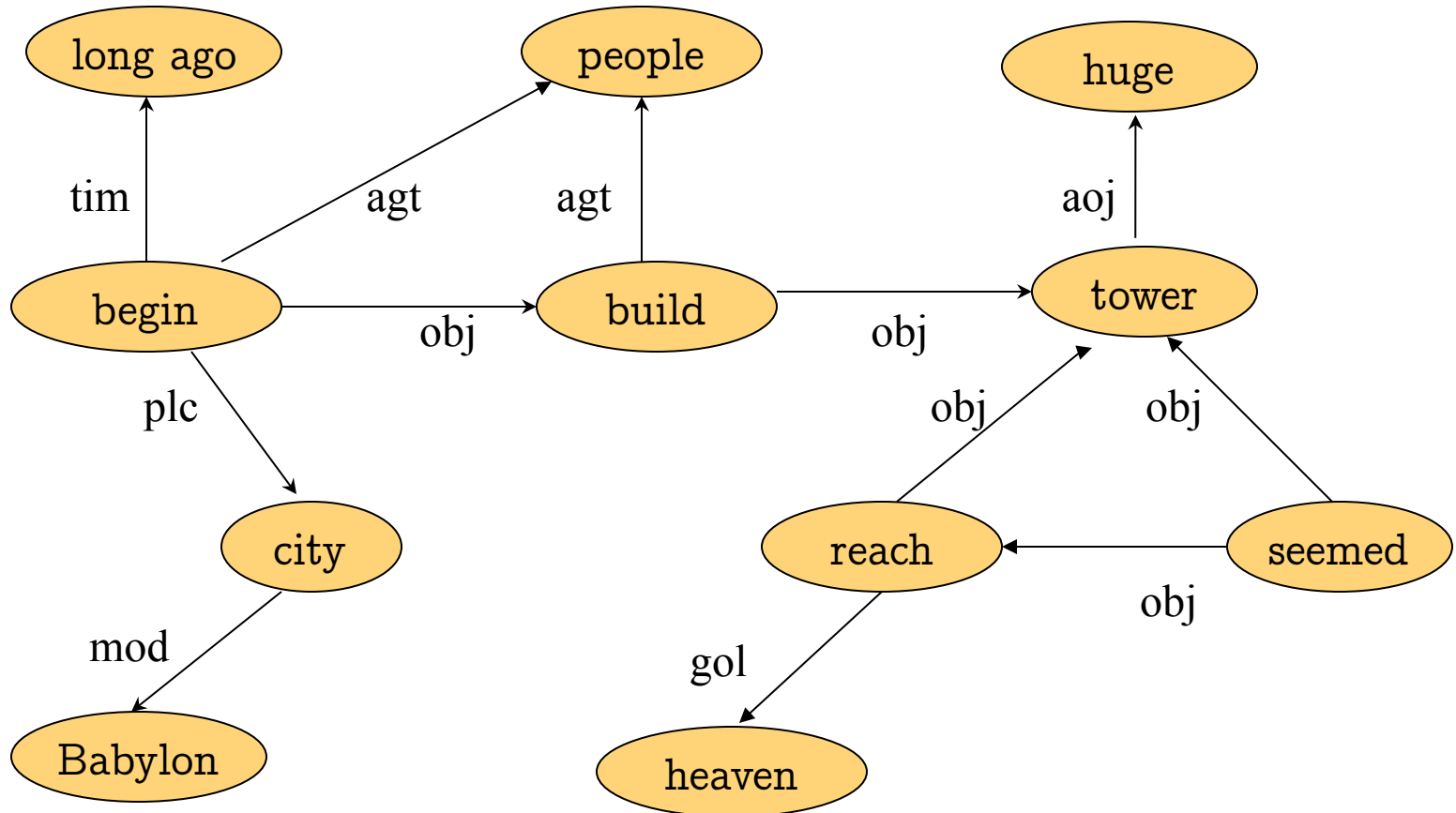
Er sah das Mädchen mit **ihrem** Teleskop

Solution: une langue non-ambigüe

Par exemple : UNL - Universal networking language
– langue formelle artificielle

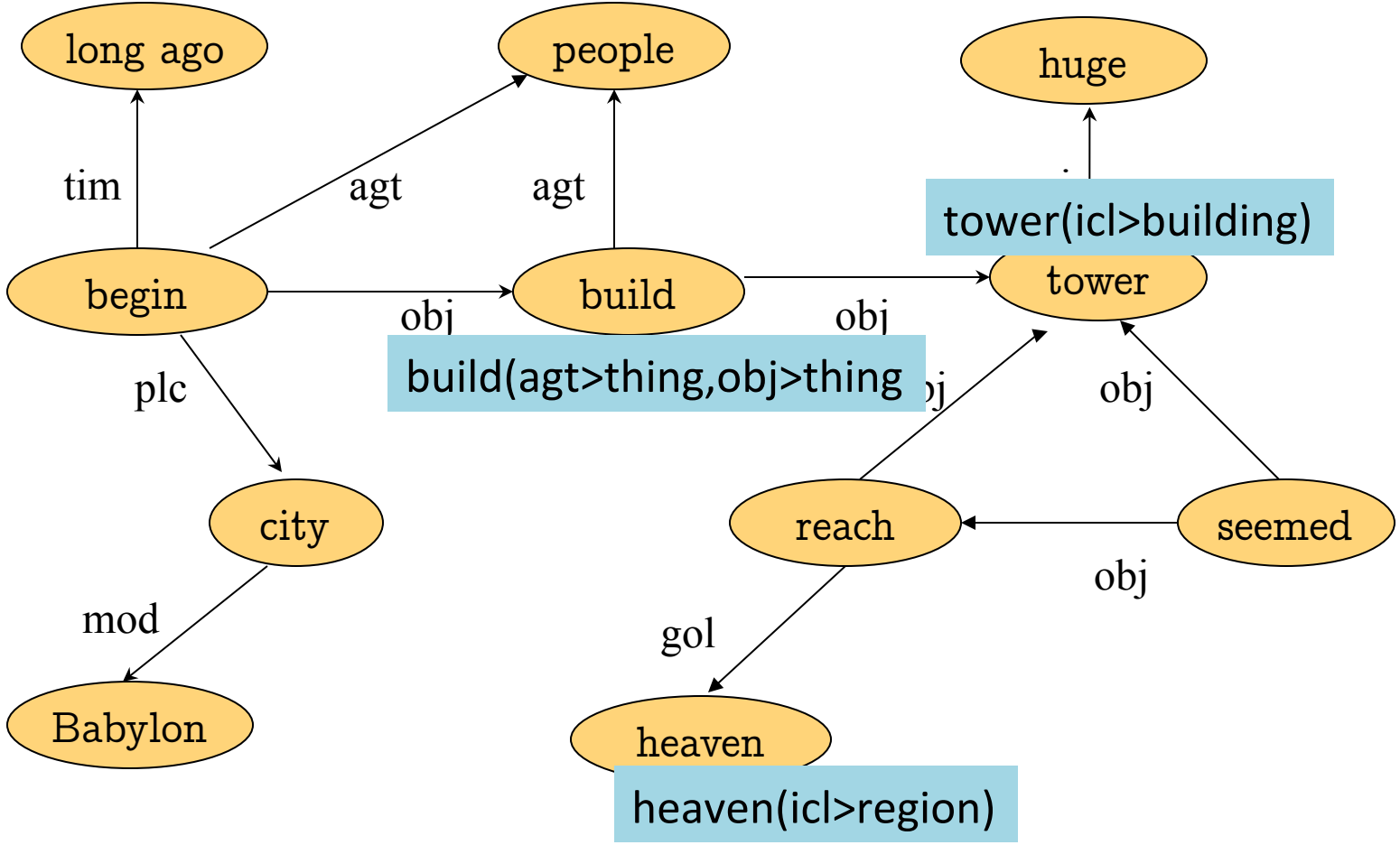


Long ago, in the city of Babylon, the people begun to build a huge tower, which seemed about to reach the heavens.



```
{unl}
tim(begin(agt>thing,obj>thing).@entry.@past,long ago(icl>ago))
mod(city(icl>region).@def,Babylon(icl>city))
plc(begin(agt>thing,obj>thing).@entry.@past,city(icl>region).@def)
agt(begin(agt>thing,obj>thing).@entry.@past,people(icl>person).@d
ef)
obj(begin(agt>thing,obj>thing).@entry.@past,build(agt>thing,obj>thi
ng)@past)
agt(build(agt>thing,obj>thing),people(icl>person).@def)
obj(build(agt>thing,obj>thing),tower(icl>building))
aoj(huge(icl>big),tower(icl>building))
aoj(seem(aoj>thing).@past,tower(icl>building))
obj(seem(aoj>thing).@past,reach(icl>come).@begin.@soon)
obj(reach(icl>come).@begin.@soon,tower(icl>building))
gol(reach(icl>come).@begin.@soon,heaven(icl>region).@def.@pl)
{/unl}
```


Mots --> Mots universels (non ambigu)



Utilisation (théorique)

Pour faire un site web multilingue

- j'écris mes pages en français
- elles sont converties, vérifiées et stockées en UNL
- quand M. 壽命長 veut les lire, le système produit des pages en chinois à partir de l'UNL

Problème: du manque d'ambiguïté

FR

Il a vu la fille avec son télescope

FR

Forcer un sens

He saw the girl with **his** telescope

ES

Él vio a la niña con su telescopio

DE

Er sah das Mädchen mit **seinem** Teleskop

Ambiguïté reconstituée. Par hasard ?

LP et recherche d'information multilingue

- Je cherche des documents en FR ou EN ou EO ou ES à propos de X Y Z.
- Technique 1: traduire les requêtes
- Problème: fusionner les réponses (ordres)
- Technique 2: "traduire" les textes dans une langue pivot "pauvre"
 - terme --> tous ses sens
 - traduction d'un document = tous les sens de tous les mots

Langages de programmation

- ça marche très bien
- l'encodage binaire des opérations est différent suivant les processeurs (Intel, ARM, PowerPC, ...)
- traduction des langages de programmation en codes binaires => multi-cibles
- idée: créer du code pour une pseudo machine
 - chaque machine traduit dans son code au moment de l'exécution
 - il n'y a qu'une version de chaque application

Pivot, connaissances et politique

YAGO

- base de connaissances tirée de Wikipédia
- entité - liens

Browse YAGO2

Search: fra ▾

<Collonge-Bellerive>

<Amadeus_III_of_Geneva>
 <Bellerive_Castle>
 <Bellevue,_Switzerland>
 <Cologne>
 <Corsier>
 <Emile_Chambon>
 <Federal_Inventory_of_Amphibian_Spawning_Areas>
 <Lake_Geneva>
 <Municipalities_of_the_canton_of_Geneva>
 <Prehistoric_pile_dwellings_around_the_Alps>
 <Prince_Sadrudin_Aga_Khan> <linksTo>
 <Versoix>

<Bellerive_Castle>
 <Bellevue,_Switzerland>
 <Buddhism>
 <Canton_of_Geneva>
 <Christian_Democratic_People's_Party_of_Switzerland>
 <Cologne>
 <Corsier>
 <English_language>
 <FDP.The_Liberals>
 <Free_Democratic_Party_of_Switzerland>
 <French_language>
 <German_language>
 <Grand_Council_of_Geneva>
 <Green_Party_of_Switzerland>
 <Hinduism>
 <Islam>
 <Judaism>
 <Lake_Geneva>
 <Liberal_Party_of_Switzerland>
 <Primary_sector_of_the_economy>
 ...

<linksTo>

<hasNumberOfPeople> "6616"^^xsd:nonNegativeInteger →

<hasWikipediaUrl> <http://en.wikipedia.org/wiki/Collonge-Bellerive>

<isPreferredMeaningOf> "Collonge-Bellerive"@eng →

"Collonge Bellerive"@eng
 "Collonge-Bellerive GE"@eng
 "Collonge-Bellerive (Geneva)"@eng
 "Collonge-Bellerive, Switzerland"@eng
 "Vesenaz"@eng
 "Vésenaz"@eng

<redirectedFrom>
 "Bellerive"@eng
 "Collonge-Bellerive"@als
 "Collonge Bellerive"@als

Politique linguistique

Wikipedia existe en de nombreuses langues

Choix : le Wikipedia sert de référence pour identifier les entités

=> un objet n'existe que s'il existe dans en.wikipedia.org

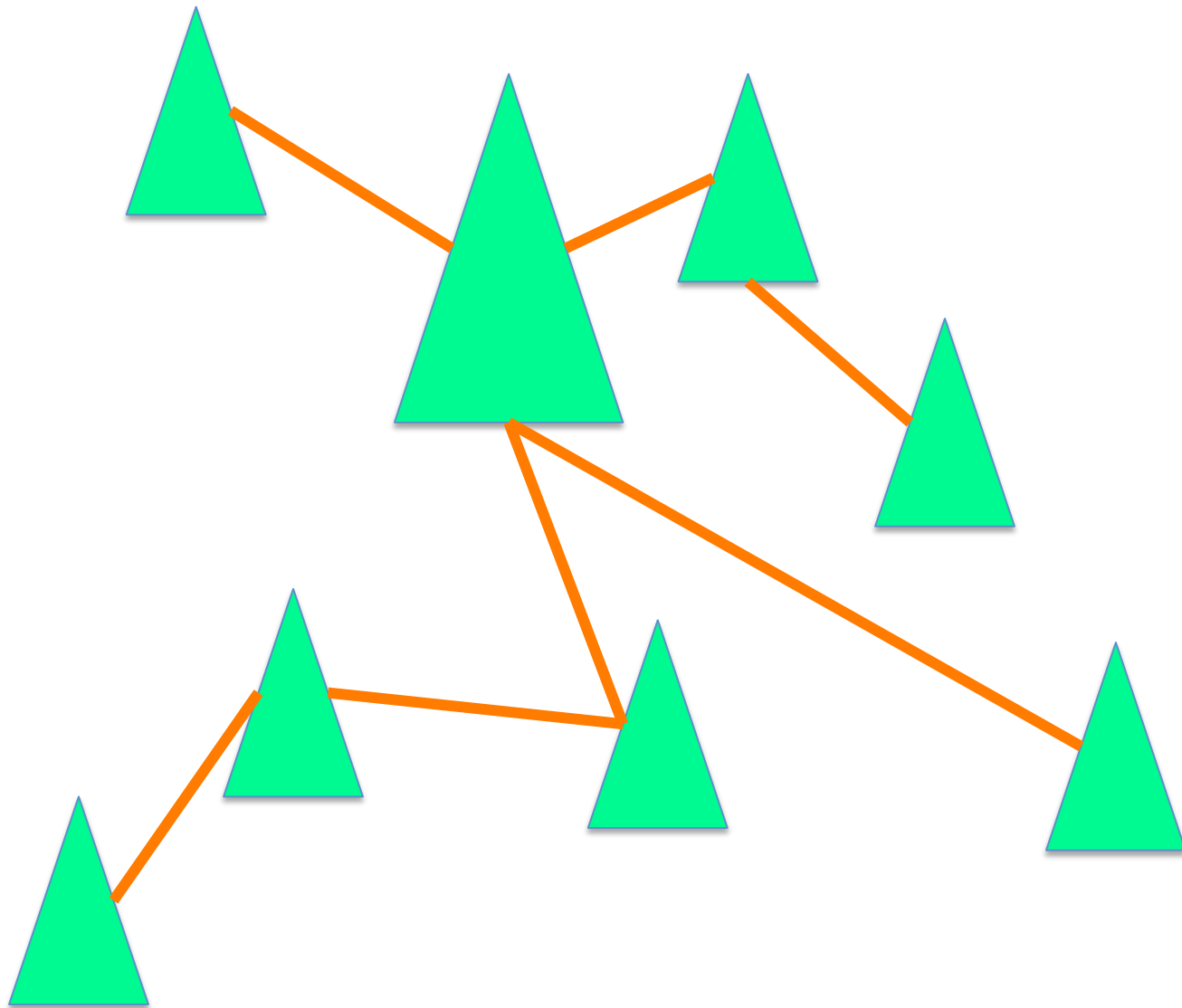
=> Vézenaz n'existe pas dans YAGO

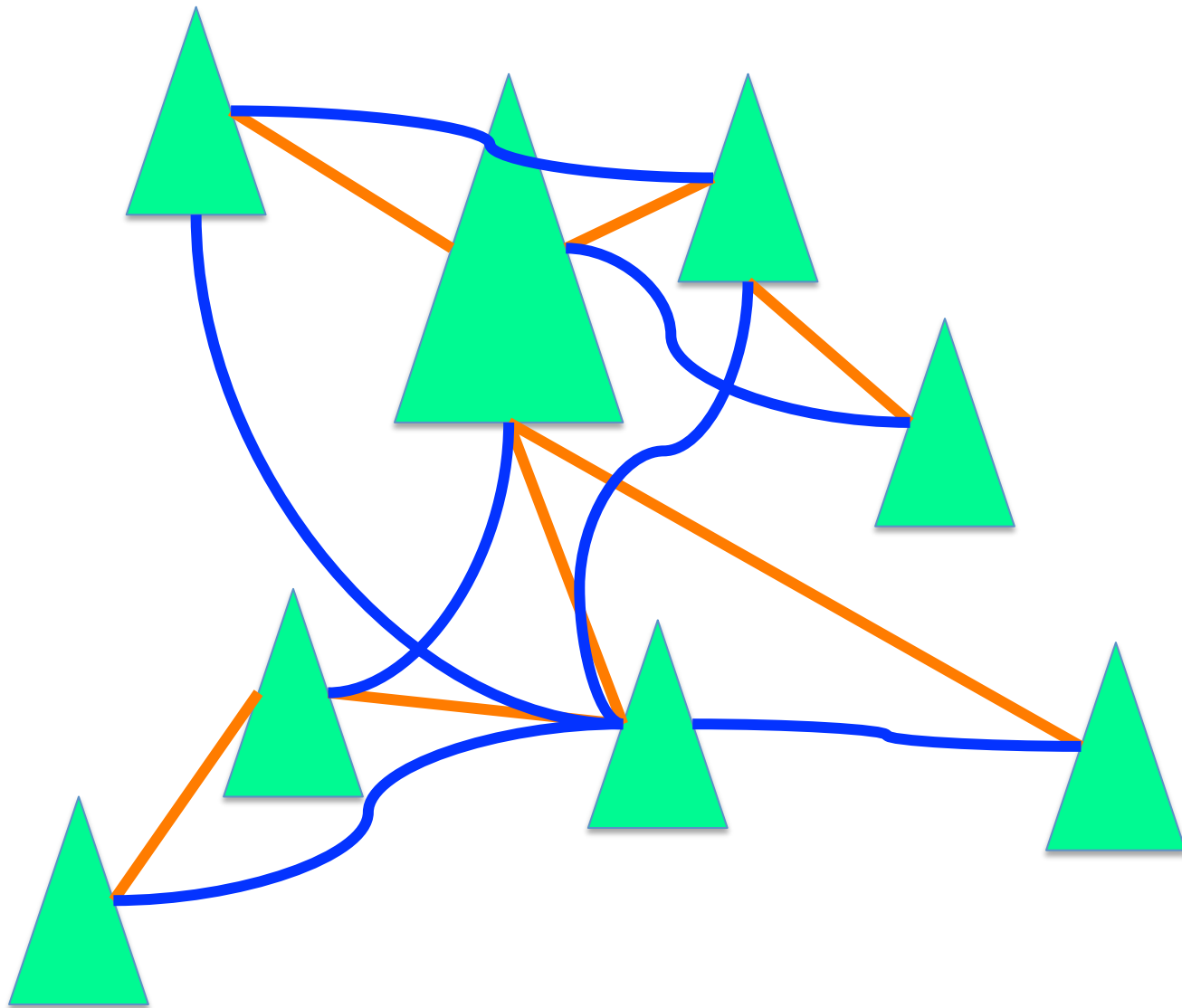
YAGO a été fait par des allemands !

=> La soumission linguistique dans toute sa splendeur

Du pivot à la composition

- Établir des correspondances entre dictionnaires, terminologies, thesaurus, ...
- Utiliser les correspondances existantes pour en calculer de nouvelles
 - si possible les meilleures
 - si possible vers les dictionnaires les plus riches
- Pas de "référence unique/centrale"





(À SUIVRE)