

## Thematic representation of general knowledge

### Larousse Thesaurus

- tree-like structure
- 873 leaf concepts
- => C873 : Vector space for modelling the thematic representation of knowledge

Fundamental, but completely fixed tool  
Can we find a "better" generative system of the vector space ?

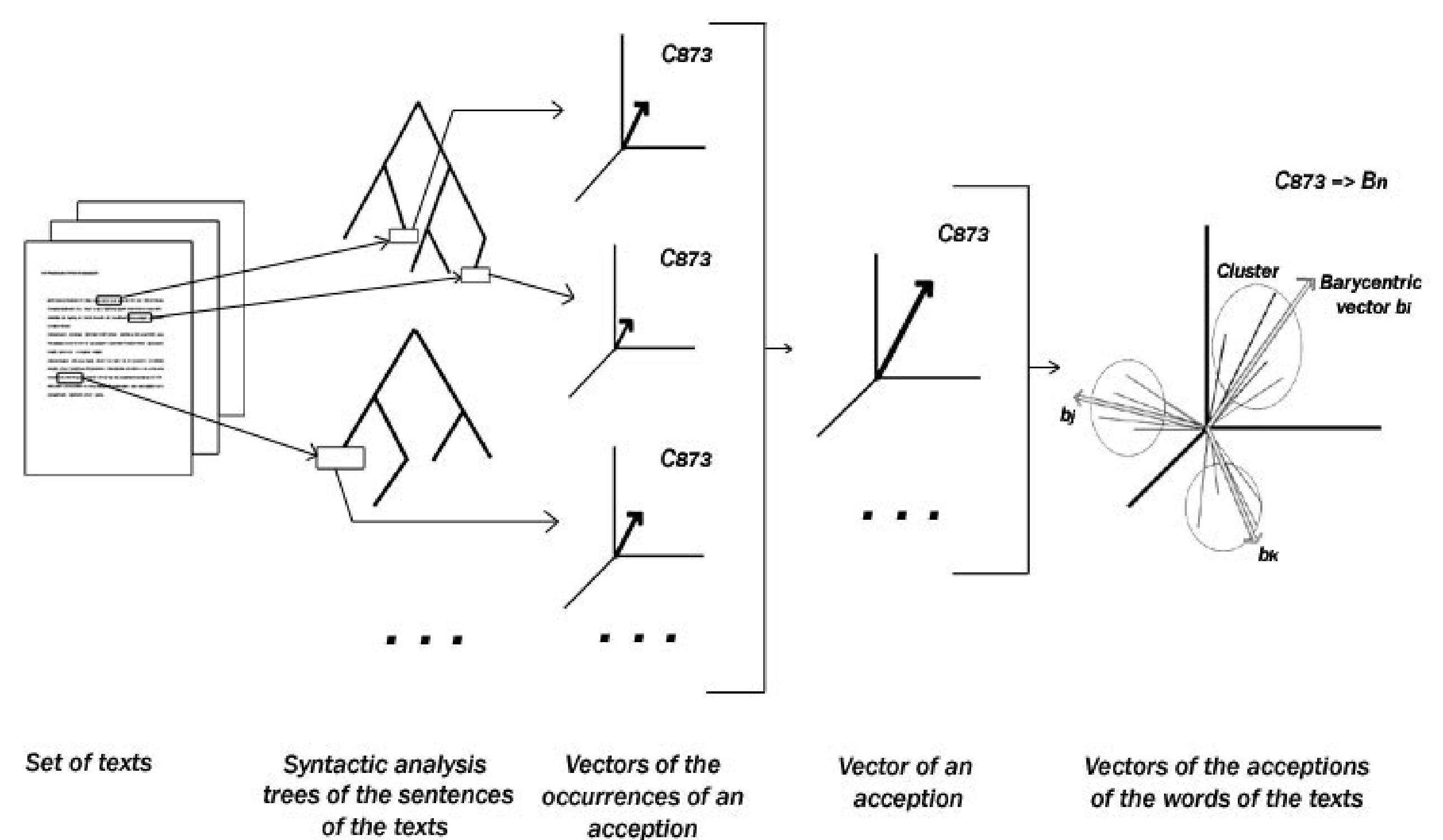
## What are the criteria of a good generative system ?

1. Extent : it is a generative system covering the whole non-specialized domain
2. Representativeness : it is a generative system close to a base
3. Evolution capacities : it is not a fixed system

## Definition and evolution of the basic notions

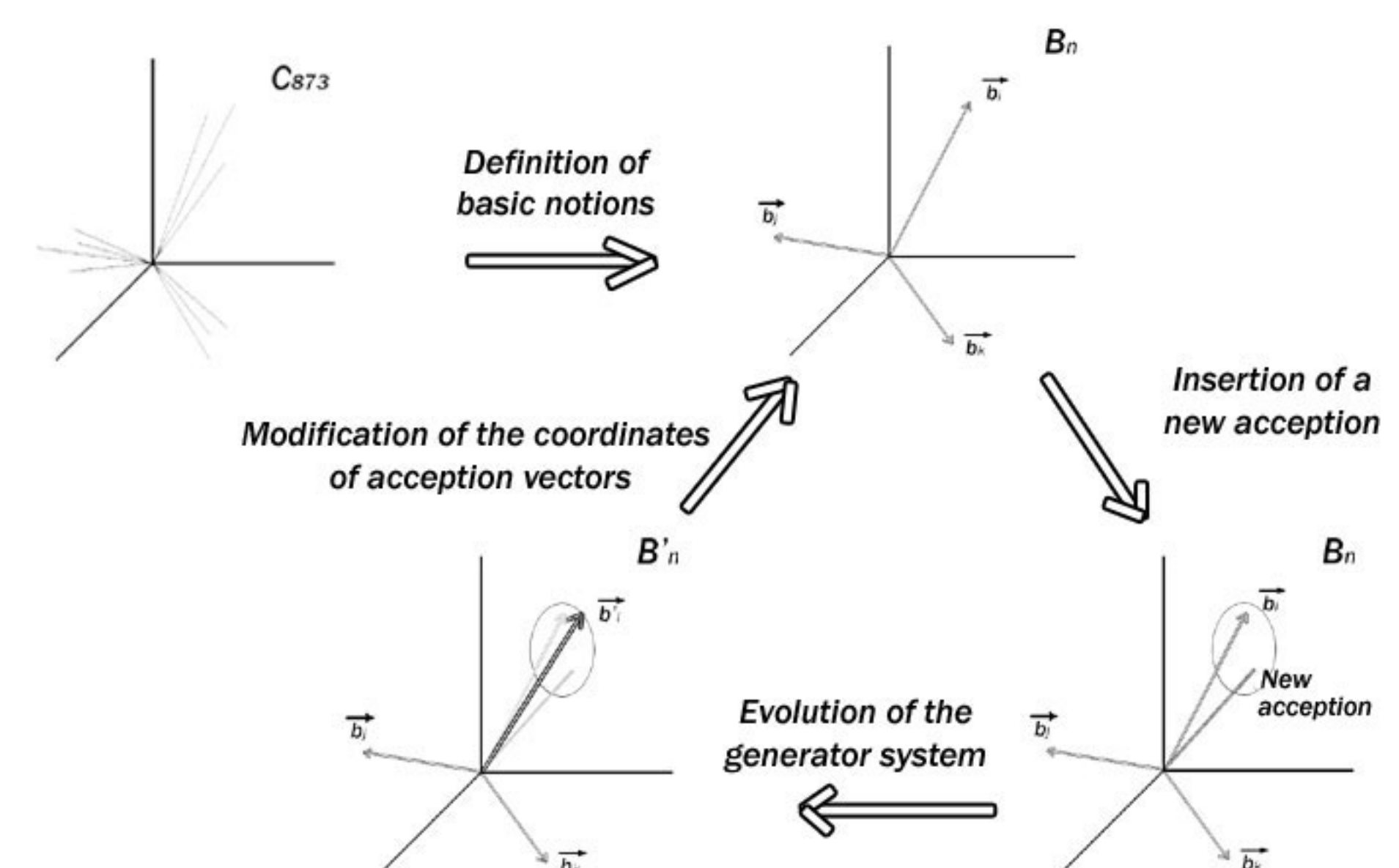
### Definition

- Studying texts  
=> Trees of morpho-syntactic analysis of the sentences
- Every occurrence of an acception gives a vector on C873.  
Summation of the vectors of the various occurrences of a same acception  
=> Vector corresponding to this acception on C873
- Grouping together vectors of various acceptions  
=> Clusters the barycenters of which define the basic notions



### Evolution

- Analysing new texts
- New acception or new occurrence of an already existing acception
- => Evolution of the clusters of vectors  
with a possible phenomenon of differentiating clusters  
or grouping them together



## Conclusion

### Evolutionary definition of the basic notions

Initial construction relies on the skeleton which constitutes the concepts of a thesaurus

Evolutionary definition totally frees itself from the fixed hierarchy of the concepts