TOWARDS A DYNAMIC COMBINATORIAL DICTIONARY: A PROPOSAL FOR INTRODUCING INTERACTIONS BETWEEN COLLOCATIONS IN AN ELECTRONIC DICTIONARY OF ENGLISH WORD COMBINATIONS

Moisés Almela, Pascual Cantos, Aquilino Sánchez
Universidad de Murcia (Spain)
LACELL Research Group
CONTENTS

1. Introduction
2. Lexical Constellations
   Theoretical background
   Techniques of description
3. Constellations of goods
   An empirical case study
4. Constellations in the dictionary
   Proposal of a new model of combinatorial dictionary:
   the Dynamic Collocation Dictionary (DCD)
5. Concluding remarks
INTRODUCTION

- Potentialities of electronic format in combinatorial lexicography:
  - Increase amount and range (variety) of contextual data
  - Facilitate interactive management of information

- This potential is underexploited. Design informed by printed dictionaries. Different medium (material format) but similar design.

- Our project: to record dependencies between collocations, not just between words (possible in electronic format, not in printed format).
The node does not exert an unlimited influence on its environment (Cantos & Sánchez, 2001).

Limitations of the concept of *lexical gravity* (Mason, 2000) = “the restriction a word imposes on the variability of its context”. Restrictions on the context of the node are not an exclusive function of the node.

Problems of *lexical gravity interference* (or *lexical gravity overlaps*).

(more about lexical constellations in Sánchez, Cantos & Almela, 2007; Almela, 2011; Almela, Cantos & Sánchez, 2011)
LEXICAL CONSTELLATIONS

- The received models of collocation = linear (no division into domains of lexical attraction).
  - Collocation = statistically significant co-occurrence.
  - Structure of a plain collocational network: shared collocation as a sufficient condition (two or more collocational bi-grams with at least one member in common).
LEXICAL CONSTELLATIONS

Lexical Constellation = collocational network hierarchically organised in two or more centres of lexical attraction.

- The context of the node is organized around two or more domains of lexical attraction.
- Shared collocation as a necessary but not sufficient condition.
Suitable for capturing inter-collocability relations (restrictions on combinations of different collocations of the same node)

Methodology: compare the influence of the node and the influence exerted by other items or structures that co-exist within the same textual window.

- More precisely: compare conditional probabilities of the type: $P(c1 | n, c2)$, where $n$ stands for the node, and $c1$ and $c2$ represent two different collocates.
Positive inter-collocability:

- **C2** is a *positive co-collocate* of **c1** if the probability of \((n,c1)\) co-occurring with **c2** is higher than the probability of the node occurring with **c2** alone. The collocation \((n,c2)\) is made more probable by the selection of **c1**.
- **C1** is a *positive co-collocate* of **c2** if the probability that \((n,c2)\) co-occurs with **c1** is higher than the probability of the node co-occurring with **c1**. The collocation \((n,c1)\) is made more probable by the selection of **c2**.
LEXICAL CONSTELLATIONS

- Negative inter-collocability:
  - $C_2$ is a *negative co-collocate* of $c_1$ if the selection of the collocation $(n,c_1)$ diminishes the probability of $(n,c_2)$. The capacity of the node for predicting the choice of $c_2$ is higher than the capacity of the collocation $(n,c_1)$ for predicting the choice of $c_2$.
  - $C_1$ is a *negative co-collocate* of $c_2$ if the selection of the collocation $(n,c_2)$ diminishes the probability of $(n,c_1)$. The capacity of the node for predicting the choice of $c_1$ is higher than the capacity of the collocation $(n,c_2)$ for predicting the choice of $c_1$. 
CONSTELLATIONS OF GOODS

- Corpus: *ukWaC* (1,565,274,190 tokens)
- Query system and tools: SketchEngine
- Queries syntactically restricted:
  - **Verb + goods**
  - **Adjective + goods**, where A+N collocation performs the semantic role of THEME (object in active construction, subject in passive construction).

Assumption: the verb is likely to exert an influence on the entire argument phrase, not only on the head. Phenomenon of *valency stratification* (collocability between predicates of the same argument is constrained).
CONSTELLATIONS OF GOODS

- Potential co-collocates: semantically related (to test hypothesis: lexical constellations can be generalized to conceptual structures):
  - *return, replace, reject* (‘consumer does not accept the goods initially bought or received’)
  - *faulty, defective, damaged* (‘flaw, imperfection’)

- Frequency threshold: 3
- Statistical filter: *logDice*
### CONSTELLATIONS OF GOODS

|                | f(v,m,n) | f(m,n) | P(m|v,n) | P(m|n) |
|----------------|----------|--------|----------|--------|
| faulty         | 35       | 354    | 2.35%    | 0.36%  |
| unwanted       | 21       | 149    | 1.41%    | 0.15%  |
| defective      | 20       | 137    | 1.34%    | 0.14%  |
| unused         | 7        | 20     | 0.47%    | 0.02%  |
| undamaged      | 6        | 10     | 0.40%    | 0.01%  |
| damaged        | 8        | 209    | 0.54%    | 0.21%  |
| non-faulty     | 4        | 11     | 0.27%    | 0.01%  |
| stolen         | 8        | 434    | 0.54%    | 0.44%  |

Table 2: Adjectival co-locates of *replace*.

Adjectival co-locates of *return*
## CONSTELLATIONS OF GOODS

Table 2: Adjectival co-collocates of *replace*.

|          | f(v,m,n) | f(m,n) | P(m|v,n) | P(m|n) |
|----------|----------|--------|----------|--------|
| faulty   | 30       | 354    | 19.11%   | 0.36%  |
| defective| 12       | 137    | 7.64%    | 0.14%  |
| damaged  | 12       | 209    | 7.64%    | 0.21%  |
| electrical| 6       | 850    | 3.82%    | 0.86%  |

Adjectival co-collocates of *replace*.

|          | f(v,m,n) | f(m,n) | P(m|v,n) | P(m|n) |
|----------|----------|--------|----------|--------|
| faulty   | 6        | 354    | 5.41%    | 0.36%  |
| defective| 3        | 137    | 2.70%    | 0.14%  |

Adjectival co-collocates of *reject*.
CONSTELLATIONS OF GOODS

|       | f(v,m,n) | f(v,n) | P(v|m,n) | P(v|n) |
|-------|----------|--------|----------|--------|
| return | 35       | 1491   | 9.89%    | 1.50%  |
| replace | 30       | 157    | 8.47%    | 0.16%  |
| receive | 19       | 913    | 5.37%    | 0.92%  |
| buy    | 17       | 1592   | 4.80%    | 1.60%  |
| reject | 6        | 111    | 1.69%    | 0.11%  |
| supply | 6        | 961    | 1.69%    | 0.97%  |
| collect | 3       | 270    | 0.85%    | 0.27%  |
| sell   | 7        | 2237   | 1.98%    | 2.25%  |

Verbal co-collocates of faulty.
## Constellations of Goods

|       | $f(v,m,n)$ | $f(v,n)$ | $P(v|m,n)$ | $P(v|n)$ |
|-------|------------|----------|------------|----------|
| return | 20         | 1491     | 14.60%     | 1.50%    |
| replace| 12         | 157      | 8.76%      | 0.16%    |
| reject | 3          | 111      | 2.19%      | 0.11%    |
| inspect| 3          | 121      | 2.19%      | 0.12%    |
| deliver| 4          | 1930     | 2.92%      | 1.94%    |

Verbal co-collocates of *defective*.

|       | $f(v,m,n)$ | $f(v,n)$ | $P(v|m,n)$ | $P(v|n)$ |
|-------|------------|----------|------------|----------|
| receive| 15         | 813      | 7.18%      | 0.82%    |
| replace| 12         | 157      | 5.74%      | 0.16%    |
| return | 8          | 1491     | 3.83%      | 1.50%    |
| inspect| 3          | 121      | 1.44%      | 0.12%    |
| deliver| 4          | 1930     | 1.91%      | 1.94%    |

Verbal co-collocates of *damaged*. 
Comparing intra-collocational and inter-collocational dependencies.

Node: *goods*. Collocate: *return*.

Grammatical class of co-collocates: adjective.
Comparing intra-collocational and inter-collocational dependencies. 
Grammatical class of co-collocates: adjective.
Comparing intra-collocational and inter-collocational dependencies.
Grammatical class of co-collocates: adjective.
Comparing intra-collocational and inter-collocational dependencies.
Grammatical class of co-collocates: verb.
Comparing intra-collocational and inter-collocational dependencies.

Grammatical class of co-collocates: verb.
CONSTELLATIONS OF GOODS

Comparing intra-collocational and inter-collocational dependencies.
Grammatical class of co-collocates: verb.
Faulty as a positive co-collocate of return in the context of goods
CONSTELLATIONS OF GOODS

- Semantic regularities: verb-noun collocations expressing ‘non-acceptance of goods’ are likely to converge with adjective-noun collocations describing goods as ‘having a flaw’.

- Semantic systematicity is also a characteristic of negative inter-collocatability: the collocations *ship/transport goods* tend to avoid the selection of modifiers describing a ‘flaw’ or ‘imperfection’.
CONSTITELLATIONS OF GOODS

- Overall tendency towards mutual inter-collocability: defective is a co-collocate of return, and conversely, return is a co-collocate of defective. The same holds for other pairs:

  - (defective, replace), (defective, reject)
  - (faulty, return), (faulty, replace)
  - (faulty, reject), (damaged, return)
  - (damaged, replace).
CONSTELLATIONS OF GOODS

DANGER
(dangerous, hazardous, etc.)

FLAW
(faulty, defective, damaged, etc.)

TRANSPORT
(transport, ship, etc.)

NON-ACCEPTANCE
(return, replace, reject, etc.)

goods

Semantically motivated interdependencies among collocates of goods
CONSTITUTIONS IN THE DICTIONARY

- Conventional collocation dictionaries provide a purely “intra-collocational” perspective. They describe dependencies between members of a collocation, not between collocations. **No specification of interactions between collocates of the same headwords.**

- This holds true for the major dictionaries of English and Spanish word combinations: *The BBI Dictionary, Oxford Collocations Dictionary, Macmillan Collocations Dictionary, Diccionario de Colocaciones del Español (DiCE), and REDES*, among others.
CONSTELLATIONS IN THE DICTIONARY

An entry from the OCD on CD-ROM
Our project: to devise a dictionary design (and eventually compile a dictionary) suitable for describing “inter-collocational” dependencies.

Focus on **positive** inter-collocability: which collocations of a headword are more likely to be activated in the same textual window?
CONSTELLATIONS IN THE DICTIONARY

Why?

1) Resource of fluency and cohesion. The word fits within a context broader than the simple collocational bi-gram.
2) Strength of patterning. Inter-collocational dependencies often stronger than intra-collocational dependencies (e.g. the dependency of the collocation defective goods on return, measured in terms of conditional probability, is ten times higher than the dependency of goods on return). The bi-gram is often weaker than the constellation.
Why?

1) Resource of fluency and cohesion.
2) Strength of patterning.
3) Accuracy in semantic description: *faulty, defective, or damage* are better represented by their verbal co-collocates (*reject, return, replace*) than by the noun (*goods*).
CONSTELLATIONS IN THE DICTIONARY

How?

1) Dynamic management of information (a *dynamic* collocation dictionary). The information presented in the entry is readjusted to the selections made by the user. **Possible only in electronic format.**

2) Progressiveness. Three stages (accessed through successive menus).

   2.1 Simple/plain collocational information (node and collocates). Semantic groups as in many collocation dictionaries (OCD, Macmillan, DiCE, etc.)

   2.2 Positive co-collocates and conceptual patterns.

   2.3 Conceptual patterns and examples.
CONSTELLATIONS IN THE DICTIONARY

Additional principles of the DCD model:

a) Compactness. Succinct format.
   - Metalinguistic information kept to a minimum. Only basic grammatical categories (Verb, Noun, Adjective, etc,) and semantic labels.
   - The structure of constellations is signalled only by means of arrows and by highlighting words in authentic examples.

b) Systematicity: subsume as much lexical information as possible under general combination rules. Surface collocations connected by semantic labels.
CONSTELLATIONS IN THE DICTIONARY

Extract from a DCD entry (first stage)

Modifier + goods

- dangerous, hazardous
- perishable vs. durable
- illegal, stolen, contraband, fake, counterfeit
- faulty, defective, damaged
- unwanted
- cheap vs. luxury
- cotton, woollen, leather,
- electronic, electrical
- agricultural, industrial
CONSTELLATIONS IN THE DICTIONARY

Verb + Modifier + goods
- dangerous, hazardous
  ⇒ TRANSPORT THINGS FROM ONE PLACE TO ANOTHER

Verb + Modifier + goods
- illegal, stolen, contraband, fake, counterfeit
  ⇒ TAKE THINGS SECRETLY TO OR FROM A PLACE

Verb + Modifier + goods
- faulty, defective, damaged
  ⇒ REFUSE TO ACCEPT THE GOODS RECEIVED OR BOUGHT
  ⇒ PROVIDE NEW GOODS

Extracts from a DCD entry (second stage)
CONSTELLATIONS IN THE DICTIONARY

Verb + Modifier + goods

⇒ REFUSE TO ACCEPT THE GOODS RECEIVED OR BOUGHT

e.g. You are not legally obliged to return faulty goods to the seller.

Defective goods were returned to the factory for rectification

Faulty or damaged goods can be returned for replacement or repair.

Usually there are no problems with rejecting faulty goods.

Why is 'notice' necessary when the buyer rejects defective goods?

Extracts from a DCD entry (third stage)
Extracts from a DCD entry (third stage)
CONCLUDING REMARKS

- There are compelling reasons to complement intra-collocational analysis with inter-collocational analysis. Just like the choice of a word restricts the choices of other words in its vicinity, the choice of a collocation constrains the choice of other collocations with the same node.
- Patterns of inter-collocational dependency are sufficiently strong to deserve lexicographical record.
- The proposal contributes to maximizing the utility of adopting an electronic format in combinatorial lexicography.
CONCLUDING REMARKS

- Compared to plain collocational analysis, constellational analysis provides a better representation of word meaning. The model offers a more exhaustive account of the connection of combinatorial and semantic properties of words.
- The Lexical Constellation model helps to abridge the distance between the collocation dictionary and the general-purpose dictionary.
Thank you for your patience
SOME REFERENCES


