KU LEUVEN

The Syntax of Idioms project
and the
Database of Dutch Dialect Idioms
(DaDDI)

"The syntax of idioms" workshop, Utrecht University, 20 January 2017

Research funded by FWO (G.A049.12N) / NWO (326-78-001)
CRISSP/KU Leuven & UiL OTS/Utrecht

CRISSP
Center for Research in Syntax,
Semantics, and Phonology

UiL OTS
UTRECHT INSTITUTE
OF LINGUISTICS

The Syntax of Idioms project

UiL OTS — **CRISSP** Center for Research in Syntax, Semantics, and Phonology

- A collaboration between researchers from Utrecht University in the Netherlands (UiL OTS) and KU Leuven (CRISSP) in Belgium

- Co-funded by the national scientific foundations of the Netherlands (NWO) and Flanders (FWO)

2

The Syntax of Idioms project

Research group:

- Norbert Corver (Utrecht)
- Jeroen van Craenenbroeck (Leuven)
- William Harwood (Leuven)
- Marko Hladnik (Utrecht)
- Sterre Leufkens (Utrecht)
- Tanja Temmerman (Leuven)

- Eefje Boef (Utrecht)
- Loes Koring (Utrecht)

3

Research Questions

- The main goal of the project – to investigate:

  1. the internal syntactic structure of idioms

  2. the external syntax of idioms, i.e. how they interact with the surrounding syntactic environment

4

The Project

- **Idioms in the literature until recently:**
  - ❖ discussion typically focuses on English
  - ❖ a relatively small set of frequently discussed idioms
  - ❖ often concerned primarily with the lexical content of idioms
  - ❖ most claims on the syntax of idioms seem to be based on informal/personal judgments

- ➢ hard to find reliable syntactic generalisations in such a small and relatively unvaried data set

5

The Project

- Fortunately, recent work has started to present systematised and extensive corpus studies (cf. i.a. Fellbaum 2006, 2007, 2015; Grégoire 2009; Horvath & Siloni 2009, 2014, 2015, 2016; Sailer 2014, 2015; the PARSEME project) as a means of demonstrating the syntactic properties of idioms.

- Still …
  - ❖ (to our knowledge) systematic study and detailed comparison on the question whether there is **cross-linguistic variation** in the syntax of idioms is quite limited;

  - ❖ moreover, little to no consideration has been given in the literature to the syntactic properties of idioms in **dialects** of one and the same language.

6

The Project

- The **empirical innovation** of this project concerns the type of data it brings to answer the research questions.

  The project:
  - ❖ focuses on functional rather than lexical material;
  - ❖ looks for the first time at idioms in a number of closely related language varieties (i.e. Dutch dialects);
  - ❖ involves a systematic collection of idiom data based on judgements by native speakers.

  - ➢ This enables us to study if and how (minor) differences in the syntax of these varieties are reflected in their idioms.

7

Database

- The collected data is made available in an online database.

- Our research group would like to thank the much needed reinforcements provided by our colleagues:

  - ➢ Alexis Dimitriadis (database developer)
  - ➢ Kristel Doreleijers (student assistant)

- We are extremely proud to present …

8

DaDDI

Database of Dutch Dialect Idioms

9

DaDDI

- The **Database of Dutch Dialect Idioms** is based on the database software for cross-linguistic (in our case cross-dialectal) research developed by Alexis Dimitriadis.

- Accessible at: http://languagelink.let.uu.nl/idioms

10

The Data

Part 1: Dictionaries

- Well over **7,000 idioms** collected from dictionaries, grammars, and idioticons of **13 dialects** spoken in the Netherlands and Flanders.

- These dialects have been chosen for
  - ➢ their geographic spread in that they represent different Dutch dialect areas;
  - ➢ certain syntactic peculiarities they exhibit (e.g. determiner drop or perfect doubling).

11

13 Locations

- Aalst
- Bree
- Brugge
- Dendermonde
- Deventer
- Drenthe
- Groesbeek
- Lede
- Leuven
- Tessenderlo
- Utrecht
- Wambeek
- Warffum

12

The Data

- Database entries for each idiom include:
  - ✓ its interpretation (in Dutch);
  - ✓ its grammatical properties such as voice, tense, aspect, modality;
  - ✓ basic information about its syntactic structure.

- This large collection of dialect idioms will be of interest to both linguists and the lay public.

13

The Data

Part 2: Fieldwork

- A second part of the data comes from formal, structured interviews with native speakers of Dutch dialects.

- Goal:
  - ➢ to collect information on the extent to which idioms are truly dependent on the functional material listed in their dictionary entries;
  - ➢ to determine the level of morphosyntactic flexibility idioms allow with regard to passivisation, topicalisation, modification, etc.

14

The Data

- **Main question:**

  Does the figurative reading remain available when the syntax of the idiom is altered?

15

Methodology

- In general: when researching a certain syntactic phenomenon in dialects, one and the same questionnaire can be used.

- This is not the case when investigating the syntax of idioms: each dialect generally has its own specific idioms.

- The structure of the questionnaire stays more or less the same across dialects, but for each dialect, a new set of idioms had to be selected:

  - ➢ 10 unique questionnaires had to be developed for 10 different dialects (5 in Flanders, 5 in the Netherlands).

16

Methodology

- **Several steps needed to develop a questionnaire**

- **1st Pilot study:**
  - ➢ Present informant with list of idioms that are of relevance for our research.
  - ➢ Informant selects the most well-known idioms from the dialect. These idioms will be used as the basis of the questionnaire.

17

Methodology

- **2nd Pilot study:**
  - ➢ Idiomatic expressions are syntactically manipulated in accordance with our research questions.
  - ➢ Informant goes through all the manipulations to ensure that the manipulations do not violate the basic grammar and vocabulary of the dialect, and that the examples remain pragmatically plausible.
  - ➢ Judgments can already be collected at this stage.
- **Roll-out of the questionnaire:**
  - ➢ Dialect speakers are asked for their judgments during formal interviews.
  - ➢ 6 speakers per dialect, on average

18

3

Example

- Dialect dictionaries show there are idioms in Dutch that involve the aspectual layer:

  Den heif   opgeiten hemmen.
  the dough up-eaten have
  ('to be the scapegoat')

- But: we need to make sure the idiomatic interpretation is genuinely ***dependent*** on such functional material with native speaker judgements.

  ➢ test whether the idiomatic interpretation is retained in the absence of the relevant functional material

  ➢ the syntactic flexibility of each idiom is likewise tested

19

Example

Q: Do you know the following expression? What does it mean?

  Den heif   opgeiten hemmen.
  the dough up-eaten have

Q: Is it possible to use this expression in the following way, with the same meaning?

a. *A at den heif   op.*   [Simple past]
   he ate the dough up
b. *Den heif   is   opgeiten geweist deir   hem.*   [Passivisation]
   The dough is   up-eaten been   by   him
c. *Den heif   eit'n   opgeiten.*   [Topicalisation]
   de dough has-he up-eaten

20

The Data

- Over **13,000 individual judgements** have been collected and are now stored in the database.

- English glosses and translations (both literal and idiomatic) have been added to this part of the data.

- The database opens up the collected data to further research by the scientific community.

21

Database demonstration

22

References

- Fellbaum, Christiane. 1993. The determiner in English idioms. In Christina Cacciari and Patrizia Tabossi (eds.*), Idioms: Processing, Structure and Interpretation. Hillsdale, New Jersey: Erlbaum. 271–296.*
- Fellbaum, Christiane. 2014. The syntax and grammar of idioms. *In Kiss, Tibor and Alexiadou, Artemis (eds.), Handbook of Syntax. Berlin: de Gruyter.* 776-802.
- Fellbaum, Christiane., Undine Kramer & Diana Stantcheva. (2003). Etwas und eins in VP-idiomen. Proceedings of the Annual Meeting of the Institut fuer Deutsche Sprache. Mannheim, Germany.
- Grégoire, Nicole. 2009. Untangling Multiword Expressions: a study on the representation and variation of Dutch multiword expressions. DoctoralDissertation, University of Utrecht.
- Horvath, Julia & Tal Siloni. 2009. Hebrew Idioms: The Organization of the Lexical Component. *Brill's Annual of Afroasiatic Languages and Linguistics* 1, 283-310.
- Horvath, Julia & Tal Siloni. 2014. Idioms: The type-sensitive storage model. Paper presented at TAU-GU Workshop on Relative Clauses, November 27-30, 2014, Goethe University, Frankfurt Sponsored by the Strategic Partnership Program (DAAD) and the DFG (FOR 1783).
- Horvath, Julia & Tal Siloni. 2015. Idioms: The type-sensitive storage model. Paper presented at BCG8L: The Grammar of Idioms, June 4-5, 2015, KU Leuven Campus Brussels.
- Horvath, Julia & Tal Siloni. 2016. Idioms: the Type-sensitive storage model. Ms., Tel Aviv University.

23

Thank You!
Questions?

24