Accounting for the Differences in the Syntactic Flexibility of Idioms

Sascha Bargmann
in collaboration with Gert Webelhuth, Manfred Sailer, and Christopher Goetze

University of Frankfurt and Goettingen

Laboratoire de Linguistique Formelle – CNRS and University of Paris-Diderot

18 March 2014
A Definition:

“An idiom is an expression larger than a word whose meaning cannot be systematically derived from meanings that the parts have when used independently of each other.”

(Huddleston and Pullum 2002, p. 273)

Prototypical Properties of Idioms:

- **phrasal**: multi-word expression
- **idiomatic**: non-literal and holistic meaning
- **lexically fixed**: none of the words can be replaced
- **syntactically fixed**: the idiom parts cannot be separated
Example of a Prototypical Idiom: *kick the bucket*

- **phrasal:** multi-word expression ✓
- **idiomatic:** non-literal and holistic meaning ✓
- **lexically fixed:** none of the words can be replaced ✓
  
  (1) a. *Tom kicked the container.*
  
  b. *Tom threw the bucket.*
  
  c. *Tom kicked a bucket.*

- **syntactically fixed:** the idiom parts cannot be separated ✓
  
  (2) a. *The bucket has been kicked.* (passive)
  
  b. *The bucket appeared to have been kicked.* (raising)
  
  c. *The bucket, Tom has kicked.* (topicalization)
  
  d. *It was the bucket that Tom kicked.* (cleft)
  
  e. *The bucket Tom kicked was astonishing.* (bare/contact relative)
  
  f. *Tom’s kicking the bucket caused great concern.* (verbal gerund)
  
  g. *Tom’s kicking of the bucket caused great concern.* (nominal gerund)
  
  h. *Which bucket did Tom kick?* (wh-interrogative)
However, not all idioms are syntactically fixed:¹

(3)  a. Tom spilled the beans.
    b. The beans were spilled by Tom.
    c. *The beans, Tom has already spilled.
    d. *These are the beans that Tom spilled.

(4)  a. Tom pulled some strings to get the job.
    b. Strings were pulled every time Tom applied for a promotion.
    c. Some strings, Tom has already pulled.
    d. The strings that Tom pulled helped Chris get the job.
    e. Tom pulled some/a lot of important strings to get you the job.

Wasow et al. (1983) and Nunberg et al. (1994):
There is a strong connection between the syntactic flexibility and the semantic decomposability of idioms.

¹ The unavailability of an idiomatic reading for (3c) and (3d) seems questionable, though.
Our Analysis

We distinguish 3 types of idioms:

1. **Syntactically Frozen** Idioms (= IPhs) like *kick the bucket*

2. **Mobile** Idioms (= ICEs)
   - (a) **Semantically and syntactically connected** idioms like *spill the beans*
   - (b) **Semantically connected** idioms like *pull strings*
Our General Analytic Strategy

1. A **syntactically frozen** (i.e. non-decomposable) idiom is listed as **one single lexical entry** (i.e. as a completely fixed tree) in the **phrasal** lexicon.

2. A **mobile** (i.e. decomposable) idiom is composed of **separate lexical entries** that are listed in the **word or phrasal** lexicon and syntactically combined in the normal way.

The **pieces of mobile idioms must be connected in the larger context containing them**.

**Our Strategy:** capture the different degrees of mobility of idiom chunks by imposing different **syntactic and/or semantic connectedness conditions** on them.
Central Assumptions concerning Syntax, Semantics, Discourse, and Idioms

Syntax:
HPSG, a constraint-based theory: no transformations, but only conditions of identity or sharing of substructure by different attributes of a common structure.

- **Passivization:**
  taken care of in the lexicon by a derivational rule (to be shown and explained later).

- **Topicalization** is done via the interplay of:
  1. the set-valued nonlocal feature SLASH,
     which takes local structures as its values, which contain the local syntactic and semantic information of an expression.
  2. a **Nonlocal Feature Principle**, 
     which, in a simplified version, states that the value of each nonlocal feature on a phrasal sign is the union of the values on the daughters.
  3. a **Complement Extraction Lexical Rule**
     (to be explained later).
Semantics:

- **Representation language:**
  predicate logic with generalized quantifiers and lambda calculus

- **Combinatorics:**
  as in Sailer (2003): a version of Flexible Montague Grammar
  with lexical type shifting and functional application at phrasal nodes

- **Semantic contribution of a sign (word, phrase, sentence, ...):**
  an expression of the semantic representation language

- **Semantic Representation (SR):**
  part of the overall representation of the sign, i.e. there can be mutual constraints
  on the syntactic form and **Semantic Representation** of a sign.
Discourse:

We assume a DRT-like architecture in which . . .

• a semantic representation of the preceding discourse is available.

• the SEMANTIC REPRESENTATION of the current sentence is still set apart from that of the preceding discourse.

• anaphoric relations have already been resolved.

Idioms:

• Each part of a mobile idiom makes a unique contribution to the SEMANTIC REPRESENTATION of the larger linguistic context.

• A part of an idiom may require the unique SEMANTIC REPRESENTATION contribution of the other part(s) of the idiom to be present in the SEMANTIC REPRESENTATION of the larger linguistic context containing it.
The **Semantic Representation** Account of Mobile Idioms

Two Case Studies:

Case Study 1: *spill the beans*

Case Study 2: *pull strings*
Case Study 1: *spill the beans*

**Descriptive generalizations** covering the empirical claims in the literature:

1. An NP of exactly the form *the beans* must be present.

2. The NP *the beans* can undergo A-movement but not A’-movement.

3. The NP *the beans* can be pronominalized.

4. Ellipsis of the verb *spill* is possible.
**spill the beans** – a Semantically and Syntactically Connected Idiom

<table>
<thead>
<tr>
<th>Lexical Entry</th>
<th>Constraints</th>
</tr>
</thead>
</table>
| \[
\begin{array}{c}
V \\
\text{spill}_{id}'
\end{array}
\] | In the **SEMANTIC REPRESENTATION**, this verb’s second argument must be specified by a term of the form \( \text{the}x[\text{beans}_{id}'(x)] \) – possibly after anaphor resolution. |

\[
\begin{array}{c}
\text{NP} \\
\lambda P. \text{the}x[\text{beans}_{id}'(x)](P(x))
\end{array}
\] | The **SYNSEM** value of this NP must occur on one of the valence lists of a word with the **SEMANTIC REPRESENTATION** \( \text{spill}_{id}' \) |

**Crucial Assumption:** The **SEMANTIC REPRESENTATIONS** in the lexical entries are contributed solely by the lexical entries themselves or pronouns/ellipsis sites licensed by them.
*S

PHON (Tom, spilled, the, lima beans)

\[
\begin{align*}
\text{category} & \quad \begin{cases}
\text{HEAD} & \quad \text{verb} \\
\text{VAL} & \quad \text{SPR} \quad \text{( )} \\
& \quad \text{COMPS} \quad \text{( )}
\end{cases} \\
\text{SR} & \quad \lambda z.\text{the}[\text{lima-beans'}(x)](\text{spill}_d'(z,x))(\text{tom}') \quad \text{error} \\
= & \quad \beta \quad \text{the}[\text{lima-beans'}(x)](\text{spill}_d'(\text{tom}',x)) \quad \text{error}
\end{align*}
\]

NP

\[
\begin{align*}
\text{phrase} & \quad \begin{cases}
\text{category} & \quad \begin{cases}
\text{HEAD} & \quad \text{noun} \\
\text{VAL} & \quad \text{SPR} \quad \text{( )} \\
& \quad \text{COMPS} \quad \text{( )}
\end{cases} \\
\text{SS} & \quad \text{LOC} \quad \text{CAT}
\end{cases} \\
\text{NP} & \quad \text{Tom}'
\end{align*}
\]

VP

\[
\begin{align*}
\text{phrase} & \quad \begin{cases}
\text{category} & \quad \begin{cases}
\text{HEAD} & \quad \text{verb} \\
\text{VAL} & \quad \text{SPR} \quad \text{( )} \\
& \quad \text{COMPS} \quad \text{( )}
\end{cases} \\
\text{SS} & \quad \text{LOC} \quad \text{CAT}
\end{cases} \\
\text{VP} & \quad \text{spilled, the, lima beans}
\end{align*}
\]

V

\[
\begin{align*}
\text{word} & \quad \begin{cases}
\text{category} & \quad \begin{cases}
\text{HEAD} & \quad \text{verb} \\
\text{VAL} & \quad \text{SPR} \quad \text{( )} \\
& \quad \text{COMPS} \quad \text{( )}
\end{cases} \\
\text{SS} & \quad \text{LOC} \quad \text{CAT}
\end{cases} \\
\text{spilled}
\end{align*}
\]

NP

\[
\begin{align*}
\text{phrase} & \quad \begin{cases}
\text{category} & \quad \begin{cases}
\text{HEAD} & \quad \text{noun} \\
\text{VAL} & \quad \text{SPR} \quad \text{( )} \\
& \quad \text{COMPS} \quad \text{( )}
\end{cases} \\
\text{SS} & \quad \text{LOC} \quad \text{CAT}
\end{cases} \\
\text{NP} & \quad \text{the, lima beans}
\end{align*}
\]
*spill* undergoing the **PASSIVE LEXICAL RULE**: 

\[
\begin{align*}
\text{word} & \quad \text{PHON} \langle \text{spill} \rangle \\
\text{category} & \quad \text{HEAD} \quad \text{VFORM base} \\
\text{SS} & \quad \text{LOC} \mid \text{CAT} \\
\text{VAL} & \quad \text{PR} \quad \langle \text{NP} \rangle \\
\text{COMPS} & \quad \langle \text{NP} \text{the beansid'} \rangle \\
\text{ARG-ST} & \quad \langle \text{NP, NP} \text{the beansid'} \rangle \\
\text{SR} & \quad \lambda y \lambda z. \text{spill}_{id}'(z, y)
\end{align*}
\]\n
\[
\downarrow
\]

\[
\begin{align*}
\text{word} & \quad \text{PHON} \langle \text{spilled} \rangle \\
\text{category} & \quad \text{HEAD} \quad \text{VFORM passive} \\
\text{SS} & \quad \text{LOC} \mid \text{CAT} \\
\text{VAL} & \quad \text{PR} \quad \langle \text{NP} \text{the beansid'} \rangle \\
\text{COMPS} & \quad \langle \rangle \\
\text{ARG-ST} & \quad \langle \text{NP} \text{the beansid'} \rangle \\
\text{SR} & \quad \lambda y \exists z. \text{spill}_{id}'(z, y)
\end{align*}
\]
the beans, were, spilled

\[ \lambda P.\text{the}\{\text{beans}\}_id'(x)(P(x)(\lambda y z.\text{spill}_id'(z, y))) \]
\[ = \beta \lambda y z.\text{spill}_id'(z, y) \]
\[ \lambda P.P(\lambda y z.\text{spill}_id'(z, y)) \]
\[ = \beta \lambda y z.\text{spill}_id'(z, y) \]
spill undergoing the **COMPLEMENT EXTRACTION LEXICAL RULE**:

\[
\begin{align*}
\text{word} & \quad \text{PHON } \langle \text{spill} \rangle \\
\text{category} & \quad \text{HEAD } \text{4 verb} \\
\text{VAL} & \quad \text{SPR } \langle \text{NP} \rangle \\
\text{COMPS} & \quad \langle \text{NP} \rangle \\
\text{LOC} | \text{CAT} & \quad \text{NLOC} | \text{INHER} | \text{SLASH} \{ \} \\
\text{ARG-ST} & \quad \langle \text{NP}, \langle \text{NP} \rangle \rangle \\
\text{SR} & \quad \lambda z \lambda y . \text{spill} id'(z, y)
\end{align*}
\[
\begin{align*}
\text{word} & \quad \text{PHON } \langle \text{spill} \rangle \\
\text{category} & \quad \text{HEAD } \text{4 verb} \\
\text{VAL} & \quad \text{SPR } \langle \text{NP} \rangle \\
\text{COMPS} & \quad \langle \text{NP} \rangle \\
\text{LOC} | \text{CAT} & \quad \text{NLOC} | \text{INHER} | \text{SLASH} \{ \} \\
\text{ARG-ST} & \quad \langle \text{NP}, \langle \text{NP} \rangle \rangle \\
\text{SR} & \quad \lambda z \lambda y . \text{spill} id'(z, y)
\end{align*}
\]
\[
\text{PHON } \langle \text{the, beans, Tom, spilled} \rangle
\]

\[
\text{PHON } \langle \text{Tom, spilled} \rangle
\]

\[
\text{PHON } \langle \text{spilled} \rangle
\]

\[
\text{PHON } \langle \text{Tom} \rangle
\]

\[
\text{PHON } \langle \text{the, beans} \rangle
\]

\[
\text{PHON } \langle \text{the, beans, Tom, spilled} \rangle
\]

\[
\text{PHON } \langle \text{Tom, spilled} \rangle
\]

\[
\text{PHON } \langle \text{spilled} \rangle
\]

\[
\text{PHON } \langle \text{Tom} \rangle
\]

\[
\text{PHON } \langle \text{the, beans} \rangle
\]
Assumptions on Pronouns

A personal pronoun is interpreted as a definite NP whose restrictor is identical to the restrictor of its antecedent:

(5)  

a. Preceding discourse: A woman$_i$ entered the room.  
   Current sentence: She$_i$ whistled.

b. Preceding discourse: $\exists x[\text{woman'}(x)](\text{enter-room'}(x))$
   Current sentence: the$_x[\text{woman'}(x)](\text{whistle'}(x))$
I was worried that the beans\(_i\) might be spilled\(_k\).

SR: \(\text{the}[\text{beans}_id'(x)](\text{worried'}(\text{speaker'}, \text{might'}(\exists z.\text{spill}_id'(z, x))))\)

but
Case Study 2: *pull strings*

*pull strings* is more flexible than *spill the beans* in several respects:

1. The determiner of the NP headed by *strings* is not frozen.

2. The NP headed by *strings* can undergo both A- and A’-movement.

3. *strings* can occur in a main clause without *pull*, if another occurrence of *strings* did cooccur with *pull* in the preceding discourse.
**pull strings** – a Semantically Connected Idiom

<table>
<thead>
<tr>
<th>Lexical Entry</th>
<th>Constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ SR \quad V \quad pull_{id}' ]</td>
<td></td>
</tr>
<tr>
<td>pull</td>
<td></td>
</tr>
<tr>
<td>In the <strong>SEMANTIC REPRESENTATION</strong>, the verb’s second argument is specified by a quantifier ( Qx ) that is restricted by ( strings_{id}'(x) ) – possibly after anaphor resolution.</td>
<td></td>
</tr>
</tbody>
</table>

| \[ SR \quad N \quad strings_{id}' \] |
| strings |
| \( strings_{id}'(x) \) restricts a quantifier \( Qx \) and either |
| a \( Qx \) binds the second argument of \( pull_{id}' \) (possibly after anaphor resolution) |
| or |
| b \( strings_{id}' \) is salient in the present discourse. |

**Crucial Assumption:** The **SEMANTIC REPRESENTATIONS** in the lexical entries are contributed solely by the lexical entries themselves or pronouns/ellipsis sites licensed by them.
\[
S \quad \text{PHON } \langle \text{Tom, pulled, (some, )strings} \rangle \\
\text{category HEAD } \verb | \text{ verb} \\
\text{VAL SPR } \langle \rangle \\
\text{COMPS } \langle \rangle \\
\text{SR } \lambda z. \exists x[\text{strings}_a'(x)](\text{pull}_d'(z,x))(\text{tom}') \\
\begin{align*}
&=_{\beta} \lambda x[\text{strings}_a'(x)](\text{pull}_d'(\text{tom}', x))
\end{align*}
\]

\[
\text{NP } \text{PHON } \langle \text{Tom} \rangle \\
\text{category HEAD } \text{noun} \\
\text{VAL SPR } \langle \rangle \\
\text{COMPS } \langle \rangle \\
\text{SR tom'} \\
\begin{align*}
\text{Tom}
\end{align*}
\]

\[
\text{V } \text{PHON } \langle \text{pulled} \rangle \\
\text{category HEAD } \text{verb} \\
\text{VAL SPR } \langle \rangle \\
\text{COMPS } \langle \rangle \\
\text{SR pull}_d' \\
\begin{align*}
&=_{\eta} \lambda y \lambda z. \text{pull}_d'(z,y) \\
&=_{\beta} \lambda Q \lambda z. Q(\lambda y. \text{pull}_d'(z,y))
\end{align*}
\]

\[
\text{NP } \text{PHON } \langle \text{(some, )strings} \rangle \\
\text{category HEAD } \text{noun} \\
\text{VAL SPR } \langle \rangle \\
\text{COMPS } \langle \rangle \\
\text{SR } \lambda P. \exists x[\text{strings}_a'(x)](P(x)) \\
\begin{align*}
\text{(some) strings}
\end{align*}
\]
\[
\begin{align*}
S &\quad \text{PHON } \langle (\text{some}, \text{strings}, \text{were}, \text{pulled}) \rangle \\
\text{SS | LOC | CAT} &\quad \text{HEAD } \text{\textbullet verb} \\
&\quad \text{VAL } \{ \} \\
&\quad \text{COMPS } \{ \} \\
\lambda P. \exists x [\text{strings}_id'(x)] (P(x)) (\lambda y \exists z. \text{pull}_id'(z, y)) \\
&= \beta \exists x [\text{strings}_id'(x)] (\lambda y \exists z. \text{pull}_id'(z, y)(x)) \\
&= \beta \exists x [\text{strings}_id'(x)] (\exists z. \text{pull}_id'(z, x)) \\
\end{align*}
\]
S

PHON ((some, strings, Tom, pulled))

SS | LOC | CAT

category
HEAD verb

VAL

SPR ( )

COMPS ( )

SR λP.∃x[strings_p(x)](P(x))(λy.pulled'(tom',y))

=β∃x[strings_p(x)](λy.pulled'(tom',y)(x))

=β∃x[strings_p(x)](pulled'(tom',x))

NP

phrase

PHON ((some, strings))

SR λP.∃x[strings_p(x)](P(x))

NP

phrase

PHON (Tom, pulled)

LOC | CAT

category
HEAD verb

VAL

SPR ( )

COMPS ( )

SR λzλy.pulled'(z,y)(tom')

=βλy.pulled'(tom',y)

VP

phrase

PHON (pulled)

LOC | CAT

category
HEAD verb

VAL

SPR ( )

COMPS ( )

SR λzλy.pulled'(z,y)}
I was worried that strings_i might be pulled_i.

SR: worried'(speaker', might'(∃x[strings_{id}’(x)](∃z.pull_{id}’(z,x))))

but
Tom and Chris graduated from law school together with roughly equal records. Tom’s uncle is a state senator, and he pulled strings to get Tom a clerkship with a state supreme court justice.

Context LF: \( \ldots \text{strings}_{id}' \ldots \)

Chris didn’t have access to any strings
Summary

1. Idioms differ from each other in their degree of syntactic flexibility.

2. Syntactically frozen idioms are treated as surface entries in the phrasal lexicon.

3. The parts of a mobile idiom each have their own lexical entry, which makes reference to some syntactic and/or semantic property of the other part(s) of the idiom.

4. Mobile idioms differ from each other in how their parts are linguistically connected.

5. We predict a hierarchy of idioms in terms of the syntactic mobility of their parts: phrasal lexical entry < syntactically connected < semantically connected.
References


