The Kicktionary
A multilingual resource of soccer language

Thomas Schmidt,
IDS Mannheim
Kicktionary

- A multilingual (German, English, French) lexical resource of soccer language
- Developed 09/2005 to 07/2006 at the Berkeley FrameNet project
- Based on corpus material (after-match reports from uefa.com and kicker.de)
- Using ideas from
  - Frame semantics, FrameNet
  - Lexical semantics, WordNet
- Research interest: electronic dictionaries (not MT, NLP, AI) for human users
Outline

1. Scenes and Frames in soccer
2. Demonstration: Kicktionary
3. (Some) problems and limitations
4. Plans / Dreams
John **bought** the sandwich from Henry for three dollars. Henry **sold** John the sandwich for three dollars. John **paid** Henry three dollars for the sandwich. The sandwich **cost** John three dollars. Henry **charged** John three dollars for the sandwich.
Scenes and Frames

- Scene: (non-linguistic) knowledge about prototypical event structures
  - with several (active) participants
  - decomposable into subevents
  - dynamic

- Frame: A set of lexical units referring to a scene
  - profiling a participant’s point of view
  - profiling a temporal point of view
Football Scene: One-on-One

PLAYER_WITH_BALL  OPPONENT  LOCATION  BALL

(1) Dremmler challenges Graziani in midfield.
(2) Graziani takes on Dremmler in midfield.

(3) Dremmler dispossesses Graziani in midfield.
(4) Graziani loses the ball to Dremmler in midfield.
Scene: ONE-ON-ONE

<table>
<thead>
<tr>
<th>CHALLENGE</th>
<th>TAKE ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>B challenges A</td>
<td>A takes on B</td>
</tr>
<tr>
<td>B makes a challenge for A</td>
<td>A is under pressure from B</td>
</tr>
<tr>
<td>B tackles A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DENY</th>
<th>BEAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>B denies A</td>
<td>A beats B</td>
</tr>
<tr>
<td>B dispossesses A</td>
<td>A nutmegis B</td>
</tr>
<tr>
<td></td>
<td>A rounds B</td>
</tr>
</tbody>
</table>
Scene: ONE-ON-ONE

<table>
<thead>
<tr>
<th>CHALLENGE</th>
<th>TAKE ON</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B</strong> stört <strong>A</strong></td>
<td><strong>A</strong> défie <strong>B</strong></td>
</tr>
<tr>
<td><strong>B</strong> bedrängt <strong>A</strong></td>
<td><strong>A</strong> est sous la pression de <strong>B</strong></td>
</tr>
<tr>
<td><strong>B</strong> attackiert <strong>A</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DENY</th>
<th>BEAT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B</strong> dépossède <strong>A</strong></td>
<td><strong>A</strong> düpiert <strong>B</strong></td>
</tr>
<tr>
<td><strong>B</strong> grätscht <strong>A</strong> ab</td>
<td><strong>A</strong> tunnelt <strong>B</strong></td>
</tr>
<tr>
<td></td>
<td><strong>A</strong> umdribbelt <strong>B</strong></td>
</tr>
</tbody>
</table>
Scenes & Frames in the Kicktionary

• 16 scenes: One-on-One, Pass, Foul, Substitution, …
• 104 frames
• 1874 lexical units
• 8164 annotated example sentences
• Ca. 3000 corpus texts
• Presented and linked on a website

⇒ DEMO
Lexical Gaps

(1) [Maradonna]\textsubscript{PASSE} spielt einen \textbf{Fehlpass}.

(2) [X]\textsubscript{OPPONENT} fängt [einen Pass von Maradonna]\textsubscript{PASS} ab.

(3) Maradonna’s pass is \textbf{intercepted}.

(4) La passe de Maradonna est \textbf{interceptée}.
Problems / Limitations

• „Static“ lexical units (persons and objects)
  – goalkeeper, defender, center forward etc.
  – bench, sideline, penalty box etc.
  – static vs. dynamic / predicative vs. denotative
  – Semantic roles? Perspective?

• Relationships within a frame
  – Intralingual (synonymy, hyperonymy etc.)
  – Interlingual (translations equivalence!)

• “I think that semantic theory must reject the suggestion that all meanings need to be described in the same terms. I think, in fact, that semantic domains are going to differ from each other according to the kind of 'definitional base' which is most appropriate to them.”
  [Fillmore 1978]
### Semantic Relations

#### Synonymy

<table>
<thead>
<tr>
<th>English</th>
<th>German</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>match – encounter</td>
<td>Strafraum – Sechzehner</td>
<td>but – cage</td>
</tr>
<tr>
<td>centre – cross</td>
<td>legen – fällen</td>
<td>frapper – tirer</td>
</tr>
</tbody>
</table>

#### Translation equivalence

<table>
<thead>
<tr>
<th>English</th>
<th>German</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>goal – Tor – but</td>
<td>schießen – shoot</td>
<td>- tirer</td>
</tr>
</tbody>
</table>

#### Hyponymy/Hypernymy

<table>
<thead>
<tr>
<th>Hyponymy</th>
<th>Hypernymy</th>
</tr>
</thead>
<tbody>
<tr>
<td>player</td>
<td>defender</td>
</tr>
<tr>
<td></td>
<td>centre-back</td>
</tr>
<tr>
<td></td>
<td>striker</td>
</tr>
</tbody>
</table>

#### Holonymy/Meronymy

<table>
<thead>
<tr>
<th>Holonymy</th>
<th>Meronymy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spielfeld</td>
<td>Hälfte</td>
</tr>
<tr>
<td></td>
<td>Sechzehner</td>
</tr>
<tr>
<td></td>
<td>Torraum</td>
</tr>
</tbody>
</table>
Plans and Dreams

- More languages
  - Spanish, Italian, Portuguese
  - Russian, Japanese, Chinese
- More corpus material
  - After match reports
  - Minute-by-minute reports
  - Spoken commentary
- More contributors
  - Kicktionary BR
  - Swedish football dictionary
  - Polish, Ukranian, …
Problems and limitations

• “FrameNet proceeds frame by frame, not word by word. This may seem a trivial point, but it isn't. Although FrameNet uses empirical data, it does not use an empirical methodology.”
  [Patrick Hanks, http://groups.yahoo.com/group/lexicographylist/]

• Kicktionary: „complete“ coverage of soccer vocabulary?

• Bottom-up methodology?
Bottom-Up vs. Top-Down

Forced a priori decisions:

- Boundaries between scenes? goal + kick-off
- Assigning a LU to (only) one frame/scene? free-kick
- Degree of abstraction?
  - Frames: „Volleying“, „Shooting“, „Move_Ball“, „Manipulate_Artefact“?
  - Frame Elements: „Volleyor“, „Shooter“, „Agent“?

Theoretical problem: no „hard“ evidence

Practical problem: Additions on lower levels (new LUs, new examples) may require restructuring on higher levels (frames), restructuring on higher levels (frames) may require changes on lower levels (example annotation)
Bottom-Up-Workflow

• Take simple decisions (with local effects) first, postpone difficult decisions (with global effects) until enough (?) empirical material has been collected
Bottom-Up-Workflow

1. Identify a Lexical Unit
2. Define it
3. Select examples from the corpus
4. Establish synonymy and translation equivalence relations
5. Annotate examples
6. Organise LUs into scenes and frames
1. Identify a Lexical Unit
header

Act of hitting the ball with the head
header

A shot carried out by hitting the ball with the head

Moments later, Vjatšeslav Zahovaiko came close to adding to Estonia's lead, sending a header from another Lindpere corner just wide. [75227 / p3]

Kristen Viikmäe drew an early save out of Akinfeev, but Russia remained dangerous and at the other end Vladimir Bystrov sent a header just over the crossbar. [75243 / p5]

After missing two chances earlier in the game, it was a case of third time lucky for Clinton Morrison, who met Damien Duff's left-wing cross to power a header past goalkeeper Nikos Panagiotou. [75266 / p4]
to volley

to receive and deliver a ball without it touching the ground

Kazakhstan captain Samat Smakov **volleyed** over the bar before Baltiev went close again just before half-time, shooting narrowly wide from the edge of the area.

First Omonia's Kekic saw his effort cleared off the line by Evgeni Iordanov, then Mujiri broke the offside trap only to shoot wide for CSKA before Kekic **volleyed** wide at the other end, all in the 52nd minute.
Kopfball

Ein mit dem Kopf ausgeführter Schuss

Acht Minuten vor Schluss der ersten Hälfte setzte Bogdani einen Kopfball am türkischen Gehäuse vorbei.

Ingemar Teevers Kopfball konnte von Ricardo pariert werden, bei der anschließenden Ecke unterlief Portugals Schlussmann allerdings den Ball, doch Costinha klärte Raio Piirajas Kopfball auf der Linie.

Slowenien hatte in der ersten Halbzeit eine große Chance zur Führung, als Gianluca Zambrotta einen Kopfball an die eigene Latte setzte.
4. Establish synonymy

Kopfball = Kopfstoß
Volley = Direktabnahme

4. Establish translation equivalence

header = Kopfstoß
header = coup de tête
Volley = volley
to flick = bugsieren (?)
4. Establish synonymy

Kopfball = Kopfstoß
Volley = Direktabnahme

4. Establish translation equivalence

header = Kopfstoß
header = coup de tête
Volley = volley

(Exploit transitivity!)
4. Establish synonymy

Kopfball = Kopfstoß
Volley = Direktabnahme

4. Establish translation equivalence

header = Kopfstoß
header = coup de tête
Volley = volley

(Exploit transitivity!)
Moments later, Vjatšeslav Zahovaiko came close to adding to Estonia's lead, sending a **header** from another Lindpere corner just **wide**. [75227 / p3]

Kristen Viikmäe drew an early save out of Akinfeev, but Russia remained dangerous and at the other end Vladimir Bystrov sent a **header just over the crossbar**. [75243 / p5]

After missing two chances earlier in the game, it was a case of third time lucky for Clinton Morrison, who met Damien Duff's left-wing cross to **power** a **header past goalkeeper Nikos Panagiotou**. [75266 / p4]
to volley

A ball that is received and delivered without touching the ground

Kazakhstan captain Samat Smakov volleyed over the bar before Baltiev went close again just before half-time, shooting narrowly wide from the edge of the area.

First Omonia's Kekic saw his effort cleared off the line by Evgeni Iordanov, then Mujiri broke the offside trap only to shoot wide for CSKA before Kekic volleyed wide at the other end, all in the 52nd minute.
Kopfball

Ein mit dem Kopf ausgeführter Schuss

Acht Minuten vor Schluss der ersten Hälfte setzte Bogdani einen Kopfbball am türkischen Gehäuse vorbei.

Ingemar Teevers Kopfbball konnte von Ricardo pariert werden, bei der anschließenden Ecke unterlief Portugals Schlussmann allerdings den Ball, doch Costinha klärte Raio Piirojas Kopfbball auf der Linie.

Slowenien hatte in der ersten Halbzeit eine große Chance zur Führung, als Gianluca Zambrotta einen Kopfbball an die eigene Latte setzte.
Vladimir Bystrov sent a header just over the crossbar.

Samat Smakov volleyed over the bar.

Zambrotta setzte einen Kopfball an die eigene Latte.
LUs and their annotations:

**Vladimir Bystrov** sent a **header just over the crossbar**.

**Samat Smakov** volleyed **over the bar**

**Zambrotta** setzte einen **Kopfball an die eigene Latte**

Frames & Scenes = Abstractions over LUs and their annotations

- LUs share a common basic meaning
- LUs have comparable perspectives on some (abstract) event
- LUs allow for comparable sets of arguments
- Arguments exhibit similar semantic relations to one another
Level 0:
• Make a comment on an existing LU,
• Make a suggestion for a new entry

Level 1:
• Integrate a new LU into the resource
• Write a definition for a LU
• Select example sentences from the corpus

Level 2:
• Establish synonymy between LUs
• Establish translation equivalence between LUs

Level 3:
• Annotate example sentences

Level 4:
• Build Scenes and Frames
• THANK YOU!
Excursion: How many LUs?

- Kicktionary: **800** German LUs
  - Systematic gaps: league vs. cup competition
  - Gaps because of small corpus size
    (< 2.000.000 words)
- „PONS-Fußballwörterbuch“ (6 languages): Around **450** German LUs
- Burkhardt: „Wörterbuch der Fußballsprache“ (monolingual): **2.200** German LUs
- „Kauderwelsch Fußball“ (7 languages): **220** German LUs
- Langenscheidt Praxiswörterbuch Fußball (3 languages): **1752** German LUs
  - Kartenvergabestrategie, anaerobe Energiebereitstellung
  - Fehlpass, Schwalbe, ausdribbeln
On 55 minutes Necati Ates was **fouled** in the box, but Fatih Tekke's penalty was saved by substitute keeper Andrey Morev, on for the injured Novikov who was rushed to hospital following the penalty incident.

---

The floodgates threatened to open. On 55 minutes Necati Ates was fouled in the box, but Fatih Tekke's penalty was saved by substitute keeper Andrey Morev, on for the injured Novikov who was rushed to hospital following the penalty incident.
Substitute Nilmar was fouled by Frank Fahrenhorst, and Berthod slotted in the resulting penalty to...

Ekranas were reduced to ten men in added time after Majus fouled in the penalty box....

After Stéphane Martine was fouled by Alfredas Skroblas, Sébastian Remy stepped up to take the penalty...

Elin Topuzakov fouled substitute Christian Wilhelmsson in the penalty area and the Arsenal FC mid...

On 55 minutes Necati Ates was fouled in the box, but Fatih Tekke's penalty was saved by substitut...

Jan Kromkamp fouled Zbigniew Grzybowski just inside the area and Jacek Dembinski sent Henk Timmer...
Kopfball

Aber van der Sar war bei einem Kopfball von Aleksandar Bajevski auf dem Posten.

Nach einem Eckball von Kapitän Mark van Bommel kam Philip Cocu zum Kopfball.

Acht Minuten vor Schluss der ersten Hälfte setzte Bogdani einen Kopfball am türkischen Gehäuse vorbei.

Ingemar Teevers Kopfball konnte von Ricardo pariert werden.

Header

Vjatšeslav Zahovaiko came close to adding to Estonia's lead, sending a header from another Lindpere corner just wide.

At the other end Vladimir Bystrov sent a header just over the crossbar.

Clinton Morrison met Damien Duff's left-wing cross to power a header past goalkeeper Nikos Panagiotou.

Kilbane went close to giving Ireland the lead glancing a header wide from Finnan's cross.
Summary

• Ideas for crowd sourcing a multilingual lexical resource of football language using FrameNet and WordNet ideas
  – In a web-based environment
  – In a bottom-up workflow
    • Practical arguments
    • Methodological arguments

• Some first considerations for implementation
If Agent is more generic than Healer, then it seems to me that there is an imbalance in the FrameNet analyses of these sentences:

- John [[Healer]] treated Mary with antibiotics.
- John [[Agent]] treated the woodwork with creosote.

The second of these sentences should have been given a frame that is more specific than simply Agent, to match the level of specificity in the first example. Or else John should be called an Agent in both cases.

To follow this to its logical extreme, shouldn't each agentive verb -- and each sense of each verb -- have its own specific type of agent? How about these examples?

- John [[stumbler]] stumbled over a root.
- John [[drinker]] drank a cup of coffee.
  
et cetera

David Frank,
http://tech.groups.yahoo.com/group/lexicographylist/message/3163
“[i]n the early papers on Frame Semantics, a distinction is drawn between scene and frame, the former being a cognitive, conceptual, or experiential entity and the latter being a linguistic one [...]. In later works, scene ceases to be used and a frame is a cognitive structuring device, parts of which are indexed by words associated with it and used in the service of understanding [...].”

Petruck, 1996

- **FrameNet (e.g. Ruppenhofer et al. 2006):**
  
  *scenario, background frame, non-lexical frame, non-perspectivized frame*

- **Kicktionary:**
  
  *scene (conceptual entity) vs. *frame* (linguistic entity)*
“[...] we do not annotate many nouns denoting artefacts and natural kinds [...]. In this area, we mostly defer to WordNet [...].”

Ruppenhofer et al. 2006: §1.1.

“I think that semantic theory must reject the suggestion that all meanings need to be described in the same terms. I think, in fact, that semantic domains are going to differ from each other according to the kind of 'definitional base' which is most appropriate to them.”

Fillmore 1978
Balanced: 10 Scenes à 10 Frames à 10 LUs
Unbalanced: 1 Scene à 240 Frames à 4 LUs

Symmetric: #frames per scene, #LUs per frame varies little
Asymmetric: #frames per scene, #LUs per frame varies much

Homogeneous: LU/Frame Element matrix is filled uniformly