4. The Kicktionary – a multilingual lexical resource of football language

Thomas Schmidt

1. Introduction

This paper presents the *Kicktionary*, an electronic multilingual (English, German, French) lexical resource of the language of football.[1] The Kicktionary was constructed predominantly on the basis of frame semantic principles, and is therefore perhaps best described as a multilingual, domain-specific FrameNet.[2] However, the objectives of the Kicktionary project are in many ways more restricted than those of the Berkeley FrameNet project. My primary goal was (and remains) to produce a lexical resource usable by *humans* for purposes of understanding, translating or otherwise paraphrasing texts in the domain of football. In contrast to much work currently being carried out by FrameNet and by related projects, the Kicktionary does thus not claim to make contributions to fields like machine translation, question answering or other sub-areas of natural language processing or artificial intelligence. By restricting the scope of research to computer-assisted lexicography for human users, I want to offer some answers to the following questions:

(1)   What types of information and what means of navigation can a dictionary structured according to frame semantic principles offer which other (printed or electronic) lexical resources do not provide?

(2)   How does a frame semantic approach support the inclusion of empirical language material (i.e. corpus examples) into a dictionary?

(3)   How does a frame semantic approach support the construction of *multilingual* lexical resources?

(4)   How does a frame semantic approach support the construction of *domain-specific* lexical resources?

(5)   What difficulties arise in a frame semantic analysis of a multilingual domain-specific vocabulary? What are the limitations of such an approach and how can they be overcome?

(6)   Does Frame Semantics have something to say about the integration of multi-medial elements into a lexical resource?

This paper is structured as follows: Section 2 gives a short review of Frame Semantics and shows how it can be applied to the domain of football. Section 3 explains how empirical evidence from a text corpus is used in that approach. Section 4 discusses aspects related to the multilinguality of the Kicktionary. Section 5 concerns difficulties and limitations of a frame semantic approach that were encountered in the analysis of football vocabulary. Section 6 introduces the concept of semantic relations which is used to overcome some of these limitations. Section 7 describes how the resulting Kicktionary is currently presented to users via a website. Finally, Section 8 provides a discussion of some broader issues relating to the use of Frame Semantics in a multilingual, domain specific lexicographic analysis.

2.   Theoretical background: Scenes and frames in football

The same reasons that make the commercial transaction event a good illustration of frame semantic principles in general (see Fillmore 1977a, b) also make football vocabulary a promising object of study for a frame semantic approach. According to Fillmore (1978: 282), a frame can be defined as "a lexical set whose members index portions or aspects of some conceptual or actional whole [i.e. a scene, T.S.]." In other words: a *frame* is a structural entity used to group linguistic expressions which share a common *perspective* on a given conceptual *scene*. Whereas a scene is de-

fined in terms of pieces of abstract (and possibly non-linguistic) knowl-
edge, the notion of a frame is concerned with the properties of concrete
linguistic means of expressing this kind of knowledge.[3]

As in a commercial transaction, the activities in a football match are
governed by a set of conventionalized rules. These rules cannot be stated
in linguistic terms alone, but they are essential to the understanding of any
linguistic way of referring to it. A football match furthermore has a clearly
definable set of actors and props taking part in it, and it is in the nature of
the game that these participants take distinct perspectives on the event
which can be reflected in different lexical choices.[4] Last but not least, a
football match as a whole is naturally decomposable into smaller sub-
events, each of which comes with its own regularities concerning the actors
and perspectives involved in it and the corresponding lexical items.

As a first example, consider the following sentences:[5]

---

3. My understanding of the terms *scene* and *frame* is based more on Fillmore's
   earlier papers about Frame Semantics than on more recent work on Frame-
   Net. Petruck (1996: 2) notes that, "[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 [...]." In the Kicktionary and in this paper I
   maintain the explicit distinction between the notions of *scene* (a conceptual
   entity) and *frame* (a linguistic entity) referred to in this quote (see also section
   8.3). The more recent literature on FrameNet (e.g., Ruppenhofer et al. 2006)
   uses terms like *scenario, background frame, non-lexical frame* and *non-perspec-
   tivized frame* all of which bear in some way on the same issues as the *scene/
   frame* distinction. I have, however, decided to work only with the latter
   because it seemed to me the most-clear cut, and also the most useful for the
   purpose of dictionary-making. In some parts of the web presentation of the
   Kicktionary, however, the term *scenario* is used. This is an accidental inconsis-
   tency – *scenario* in this context is to be understood in precisely the same sense
   as *scene*.
4. *Actors* and *props* are terms used by Fillmore in his earlier papers. For
   instance, the commercial transaction event has a buyer and a seller as actors,
   and the goods and the money exchanged as props (Fillmore 1978). When
   actual scenes and frames are defined, actors and props are represented as FEs
   (see below).
5. These and all following examples are based on attested corpus examples from
   the corpus described in section 3, but have been shortened and/or simplified
   for the purpose of this paper.

(1)   a.   [Zahovaiko]~OPPONENT_PLAYER~ *challenged*
[Manou Schauls]~PLAYER_WITH_BALL~ [in the penalty area]~AREA~.

   b.   [He]~PLAYER_WITH_BALL~ turned inside to *take on*
[Roma]~OPPONENT_PLAYER~ and finish with his left foot from
close range.

   c.   [Hector Font]~PLAYER_WITH_BALL~ tried to *nutmeg*[6]
[Ioannis Skopelitis]~OPPONENT_PLAYER~.

   d.   [Ronaldo]~OPPONENT_PLAYER~ *dispossessed*
[Wisla goalkeeper Radoslaw Majdan]~PLAYER_WITH_BALL~
[on the edge of the box]~AREA~.

The lexical units (henceforth: LUs) *challenge*, *take on*, *nutmeg* and *dispossess* in these examples all evoke the same scene, namely a one-on-one situation in which a fixed set of actors and props (henceforth: frame elements – FEs[7]) takes part: a player in possession of the ball (PLAYER_WITH_BALL) is attacked by an opponent (OPPONENT_PLAYER) at some location (AREA) on the field.[8] Each example, however, imposes a somewhat different perspective on that scene. Thus, in (1a) and (1b), the temporal focus is on the event itself, while (1c) and (1d) relate the event from the perspective of its outcome. Similarly, (1a) and (1d) foreground the point of view of the opponent player, while (1b) and (1c) focus on the player in possession of the ball. This way of relating different LUs to one another

---

6. To *nutmeg* an opponent means to beat him in a one-on-one situation by playing the ball through his legs, rounding him, and collecting the ball again behind his back.

7. Given the explicit distinction between scenes and frames explained above, it would be more consistent to call these actors and props *Scene Elements*, since they are conceptual, rather than linguistic entities and remain constant across different frames belonging to the same scene. However, as this is bound to create confusion among readers who are familiar with FrameNet terminology, I decided to use the term *Frame Element* in this paper. Here and in the remainder of the paper, the following conventions are used: LUs are written in italics (*nutmeg*), FEs are written in small capitals (PLAYER_WITH_BALL), the names of frames are written in an equidistant font (`Challenge`), and the names of scenes are in bold face (**One-on-One**).

8. Due to space limitations it is not always possible to provide full descriptions of the frames, scenes, and parts thereof. Please point your internet browser to [http://www.kicktionary.de] to get access to complete descriptions.

by associating them with the same scene and differentiating them according to the perspective they impose on that scene is useful for structuring a large number of vocabulary items. Thus, LUs like *beat*, *outstrip* or *sidestep* have similar properties with respect to this scene-and-perspective distinction as the verb *nutmeg*. These LUs are therefore all assigned to the same frame Beat. Likewise, the verbal LU *tackle* and the nominal LU *sliding tackle* share their perspective on the **One-on-one** scene with the verb *challenge*. These LUs are therefore all assigned to the same frame Challenge.

A similar scenes-and-frames analysis can be carried out for many other areas of football vocabulary. For example, the **Foul** scene refers to a prototypical sequence of events as in the following description:

1. A player (the OFFENDER) or a whole team (the OFFENDER_TEAM) commits some kind of infringement of the laws of the game, typically (but not necessarily) involving a player of the opponent team (the OFFENDED_PLAYER), e.g., a foul, an offside position or a handball.
2. The REFEREE reacts to this infringement (the OFFENSE), by imposing a sanction on the OFFENDER (e.g. cautioning him) and/or by awarding a COMPENSATION (e.g., a penalty kick) to the opponent team (the OFFENDED_TEAM).

The following set of sentences demonstrates what different lexical choices can be made to foreground one aspect of this scene and background, or even omit others:

(2) a. [Costinha]$_\text{OFFENDER}$ *tripped* [Ignashevich]$_\text{OFFENDED\_PLAYER}$.

   b. [The referee]$_\text{REFEREE}$ *awarded* [a penalty]$_\text{COMPENSATION}$ [to CSKA Moscow]$_\text{OFFENDED\_TEAM}$.

   c. [Ignashevich]$_\text{OFFENDED\_PLAYER}$ *won* [a penalty]$_\text{COMPENSATION}$ [for CSKA Moscow]$_\text{OFFENDED\_TEAM}$.

   d. [Costinha]$_\text{OFFENDER}$ *conceded* [a penalty]$_\text{COMPENSATION}$ [by tripping Ignashevich]$_\text{OFFENSE}$.

   e. [The referee]$_\text{REFEREE}$ *cautioned* [Costinha]$_\text{OFFENDER}$ [for his foul on Ignashevich]$_\text{OFFENSE}$.

Further examples of prototypical events around which football scenes are constructed include shots, passes, goals, substitutions or the match as a whole. With this overview, I now turn to a discussion of the workflow that underlies the Kicktionary project.

3. Workflow

Once a given LU is identified as belonging to a specific scene and frame, example sentences can be searched for in a corpus and annotated according to that analysis.[9] This involves identifying the actual form of an LU as well as the realizations of its FEs (see the examples 1 and 2 above).

More than half of the LUs in the Kicktionary are nominal expressions, which have been analyzed and annotated using the same principles used for verbal LUs. The following sentences illustrate different annotations for the (compound) noun *overhead kick*, which is part of the Shoot frame.

(3) a. [Davide Furlan's]$_{SHOOTER}$ *overhead kick* found Francesco Ruopolo on the penalty spot.

  b. [Francesco Ruopolo]$_{SHOOTER}$ answered by <u>attempting</u> an *overhead kick* at the opposite end.

In (3a), the FE SHOOTER is integrated as a specifier into the noun phrase which has the LU as its head. In (3b), a support verb *attempt* connects the LU with its FE syntactically. Support verbs are systematically recorded in this way for all nominal LUs. The far less frequently occurring adjectival or adverbial LUs are treated in a similar fashion as example (4) illustrates for the LU *ahead* in the Lead frame:

(4) By now Celtic were aware that [Shakhtar]$_{LEADER}$ <u>were</u> [2-0]$_{SCORE}$ *ahead* [against Barcelona]$_{TRAILER}$ in the Ukraine.

Having discussed how different types of English LUs are annotated as part of the workflow, I now turn to a discussion of how LUs from different languages are treated in the Kicktionary.

---

9. The corpus used for the construction of the Kicktionary consists of English, French and German football match reports taken from the website of the Union of European Football Associations (UEFA, www.uefa.com). For each language, about 500 such texts, amounting to roughly 250,000 words, were used. The German part of the corpus was supplemented with about 1,000 similar reports (approximately 700,000 words) from the website of the journal Kicker (http://www.kicker.de) and with a small number of transcriptions of live commentary from German radio (approximately 10,000 words).

4. Interlingual scenes, multilingual frames

The question of how to link lexical information from different languages is one major issue in the creation of multilingual lexical resources. The Kicktionary project suggests that scenes and frames are useful for this purpose since they are by definition independent of specific languages. It thus seems plausible to assume that, at least as far as the domain of football is concerned, a native speaker of English has a very similar abstract knowledge of prototypical events in that domain as a native speaker of German or French (provided, of course, that they have comparable levels of knowledge about football). Given this state of affairs, it should be possible to use a scenes-and-frames analysis of a given domain in one language as a type of language-neutral structural backbone of a multilingual resource. This is comparable to what Boas (2005a: 457) describes as "stripping the FrameNet database of its English-specific lexical descriptions" and then "re-populating the database with non-English lexical descriptions". One major difference to Boas' (2005a: 457) proposals is that in the Kicktionary workflow frames are "populated" more or less simultaneously with lexical material from English, German, and French, as it was planned as a multilingual resource from the outset. The result is a scenes-and-frames hierarchy which can be applied in principle across individual languages, and frames which can contain LUs from different languages.

Between the LUs of a given frame or scene, various types of cross-linguistic correspondences and divergences can be found, and a frame semantic analysis helps to classify and explain these relationships.

First, consider cases in which a LU and its translation equivalent, if it exists, are members of the same frame. In the simplest case, this is a pair of LUs in two languages whose meanings, parts of speech, and argument structure are largely identical, such as with the English LU *nutmeg* and its German counterpart *tunneln* ('to (make a) tunnel'[10]) – both part of the Beat frame in the **One-On-One** scene:

(5) a. [Hector Font]~PLAYER_WITH_BALL~ tried to *nutmeg* [Ioannis Skopelitis]~OPPONENT_PLAYER~.

   b. [Ailton]~PLAYER_WITH_BALL~ *tunnelte* [Chris]~OPPONENT_PLAYER~ und spielte so Klasnic frei.

---

10. Here and in what follows, the English glosses for French or German LUs attempt to capture the literal (i.e., non-metaphoric) meaning of the item in question.

Second, consider cases where two LUs share the same semantic characteristics and argument structures, but differ in their part of speech. They are nevertheless assigned to the same frame, as the nominal French LU *petit pont* ('little bridge') in (6), which is arguably the best translation of the English verb *nutmeg* in the Beat frame, illustrates.

(6)   [Bastian Schweinsteiger]$_{PLAYER\_WITH\_BALL}$ manquait le cadre après
       avoir réussi un *petit pont* [sur William Gallas]$_{OPPONENT\_PLAYER}$.

Next, there are also cases of translation equivalence where the meaning and part of speech of two LUs are identical, but the grammatical properties of the LUs differ in some aspect. In such cases, the annotated examples are useful for detecting these differences. Thus, the sentences in (7) indicate that the English LU *play* in the Match frame (in the **Match** scene) and its German equivalent *spielen* behave differently with respect to number agreement (TEAM1 is plural in English, singular in German), and may differ with respect to the form of their object (direct object in English, prepositional object in German):

(7)   a.   On that day [Northern Ireland]$_{TEAM1}$ *play* [England]$_{TEAM2}$
             [at Old Trafford]$_{MATCH\_LOCATION}$.

       b.   [Wales]$_{TEAM1}$ *spielt* [in Cardiff]$_{MATCH\_LOCATION}$
             [gegen Nordirland]$_{TEAM2}$.

In those cases where no direct translation equivalent for a given LU exists, the information encoded in the scenes-and-frames structure of the Kicktionary can be helpful in identifying potential paraphrases in the target language. For example, (8) is an annotated example of the French LU *coup du sombrero* ('sombrero move'), which means (the act of) getting past an opponent by lobbing the ball over him, rounding him and retrieving the ball behind his back.

(8)   [Ronaldinho]$_{PLAYER\_WITH\_BALL}$ [lui]$_{OPPONENT\_PLAYER}$ faisait le *coup du sombrero*.

Neither English nor German offer a lexicalized way of expressing the same concept. The available alternatives include using a complex paraphrase like the one given in the previous paragraph, or using an LU that expresses the same general idea, but is less specific than the source expression such as a verbal hypernym. If such LUs exist, they will again be members of the same frame. For (8), the relevant frame Beat could, for instance, provide the user with LUs such as the English verb *round* or the

German verb *ausspielen* ('out-play'), both of which are fairly adequate (if less specific) translations of *(faire le) coup du sombrero.*

In other cases, it is possible to compensate for a missing translation equivalent by using another member of the corresponding frame together with an appropriate FE. For instance, German does not have a LU expressing the same idea as the English *side-foot,* i.e., to shoot with the side of the foot:

(9)   [He]_{SHOOTER} calmly rounded Marshall before *side-footing* [the ball]_{BALL} [into the net]_{TARGET}.

However, the frame Shot, which contains the LU *side-foot*, offers several German verbs whose annotated examples indicate that and how a FE PART_OF_BODY can be used with them. Via the frame assignment, a user of the resource can thus discover a way of paraphrasing (9) by employing, for instance, the German LU *bugsieren*:

(10)   [Er]_{SHOOTER} spielte Marshall aus und *bugsierte* [den Ball]_{BALL} [mit dem Innenrist]_{PART_OF_BODY} [ins Netz]_{TARGET}.

There are also cases where a particular frame is language-specific, i.e., where one language offers a way of linguistically expressing a certain perspective on a given scene, while another language does not. While these are not very common in the football domain, (11) shows a particular usage of *take on*, which profiles a one-on-one situation from the perspective of the player with the ball:

(11)   [Maris Verpakovskis]_{PLAYER_WITH_BALL} *took on* and beat [centre-half Nowotny]_{OPPONENT_PLAYER} before squaring the ball for Kleber.

Whereas French offers *défier* ('defy') as a good direct translation equivalent, German does not have a lexicalized means of expressing the same perspective on a one-on-one scene. In other words, the corresponding frame Take_On contains only English and French, but no German LUs. In order to arrive at an adequate German translation of (11), the Kicktionary user will consult other frames belonging to the same scene. The description of the corresponding scene **One-On-One**, for instance, reveals that LUs in the frame Challenge take the opposite perspective of those in the frame Take_On. They relate a one-on-one situation from the perspective of the attacking player. Among the German LUs in this frame is the verb *angreifen* ('attack'), which, if passivized, adequately paraphrases (11) as shown in (12):

(12)   [Maris Verpakovskis]_{PLAYER_WITH_BALL} *wurde*
       [von Innenverteidiger Nowotny]_{OPPONENT_PLAYER} *angegriffen*,
       umdribbelte ihn und spielte einen Querpass auf Kleber.

Alternatively, the frame `One-On-One` contains LUs taking a neutral perspective on the same scene. The German noun *Zweikampf* ('two-fight') is a member of this frame and provides another means of paraphrasing (11) as shown in (13):

(13)   [Maris Verpakovskis und Innenverteidiger Nowotny]_{PLAYERS}
       lieferten sich einen *Zweikampf*. Verpakovskis setzte sich durch
       und spielte einen Querpass auf Kleber.

5.   Difficulties and limitations of the scenes-and-frames analysis

As described in Section 2, lexical items from the football domain often lend themselves very naturally to a frame semantic approach. However, as with all lexicographic work, there are also cases where an unequivocal analysis of a given lexical item becomes more difficult.

Nouns whose main function is to denote persons and objects (like *goalkeeper, substitute, byline, penalty area*) rather than to describe processes or activities (like most LUs exemplified in the previous sections) constitute a class of words that are especially difficult to characterize. In this case the concept of scenes and frames loses a lot of its intuitiveness.[11] The notion of perspective, needed to characterize the relationship between a scene and the frames that belong to it, is therefore less easily applicable in "static" scenes (e.g. **Actors** or **Field**) which were introduced to the Kicktionary to accommodate such words.

Another type of difficulty arises from the lack of clear boundaries between the scenes of a football match. For instance, the fact that the match is restarted by a kick-off after a goal has been scored may be an argument in favor of including the LU *kick-off* (as a member of an appropriate frame) in the **Goal** scene. At the same time, an argument against such an analysis is the fact that a kick-off is carried out at a different loca-

---

11. This is also likely to be one of the reasons for the general language FrameNet to neglect such words: "[...] 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). It is worth noting, however, that, at least in the football domain, such nouns constitute a significant portion (more than 25%) of the overall vocabulary.

tion on the field, and by actors who do not have a direct connection to any FE of the rest of the **Goal** scene. In this particular case, I decided not to treat the kick-off event as a part of the **Goal** scene, mainly because it would have meant the introduction of a new FE to the scene exclusively for the description of this one LU. This decision, however, is arguably based more on pragmatic considerations (e.g., economy of design) than on purely linguistic principles.

A similar problem was encountered in the assignment of the LU *free-kick* to its "correct" frame and scene. Since a free-kick is by necessity preceded by an infringement of the laws of the game and a subsequent referee intervention, it seems plausible to regard it as belonging to a final stage of the **Foul** scene (see above). However, as with the LU *kick-off*, the FEs used with the LU *free-kick* are different from the FEs of the rest of the scene – the player who executes a free-kick is not necessarily identical to the OFFENDED_PLAYER, and the TARGET or the RECIPIENT of a free-kick are two further FEs that do not figure anywhere else in the **Foul** scene:

(14)   a.   [Sonck]EXECUTING_PLAYER sent a *free-kick*
             [into the top right corner]TARGET [from 20 metres]SOURCE.

       b.   [Anton Naumov]EXECUTING_PLAYER floated a *free-kick* [into the penalty box]TARGET [for defender Tomas Mikuckis]RECIPIENT.

In fact, (14a) and (14b) demonstrate that, instead of emphasizing its role as a compensation for a foul, a free-kick might equally well be analyzed as a special type of shot or pass and thus be assigned to an appropriate frame in the **Shot** or **Pass** scene, respectively. In this case, I chose the first alternative (i.e. assign *free-kick* to a frame Set-Piece in the **Foul** scene). Again, this was not based on an irrefutable linguistic analysis, but rather on pragmatic considerations about which analysis would result in the most economic data structure and thus in an organization of the lexicon which is maximally transparent to a user.

Another kind of difficulty arose with the definition and delineation of frames within a scene. Thus, the scene **Shot** must provide appropriate frames to accommodate both LUs like *shot* and *shoot*, as well as LUs describing an opponent's interaction with a shot. The verbs *block* and *fist* are examples of such LUs:

(15)   a.   [Jon Dahl Tomasson's point-blank shot]SHOT *was blocked*
             [by Greek defender Kostas Katsouranis]INTERVENING_PLAYER.

       b.   [Casillas]GOALKEEPER *fisted* [away]INTERVENTION_TARGET
             [Candela's deflected shot]SHOT.

There are good reasons to include these two LUs in the same frame, or alternatively, to create two separate frames for them. On the one hand, the label GOALKEEPER in (15b) is only a more specific label for the INTERVENING_PLAYER of (15a). Seen from a sufficiently abstract point of view, their role in and perspective on the scene is the same, hence the two verbs could go into the same frame. On the other hand, it may be argued that a goalkeeper's interaction with a shot is sufficiently distinct from an arbitrary player's interaction to regard the two as different possible outcomes of the same event, and hence to make two different frames for the LUs in question. Again, the actual decision was taken on the basis of pragmatic considerations: since there was a large number of LUs both for describing the more general interventions of an arbitrary player (e.g., *deflect, clear, turn*) and for describing the more specific interventions of a goalkeeper (e.g. *parry, punch, palm*), I decided to have two separate frames (`Intervention` and `Save`, respectively) and to state their close relatedness in the verbal description of the corresponding **Shot** scene.

6.   Synonymy, translation equivalence and other semantic relations

So far, the scene-and-frame hierarchy does not include information about basic semantic relations. Consider, for example, the frame `Shot`, which contains the following English, German, and French LUs, among many others:

(16)   a.   *shot, drive, thunderbolt, volley, bicycle kick, overhead kick, header, diving header*

   b.   *Schuss, Torschuss, Hammer, Volley, Direktabnahme, Fallrückzieher, Kopfball, Kopfstoß, Flugkopfball, Kopfballtorpedo*

   c.   *tir, frappe, boulet de canon, vollée, retourné, tête, coup de tête, tête plongeante*

Grouping these nouns together is justified by an analysis that assumes that they all impose the same perspective (namely the shooter's) on the same prototypical scene (namely a shot). While a scene-and-frames analysis thus captures an important commonality between these words on a relatively abstract semantic level, it does not provide information about a number of other, more basic, semantic relations between them such as the following:

1. Synonymy. The LUs *Kopfball* ('head ball') and *Kopfstoß* ('head kick') are synonymous, as are *bicycle kick* and *overhead kick*, as well as *tête* ('head') and *coup de tête* ('head kick'). Whereas synonymy in these cases is also reflected by a morphological component common to both members of the pairs, other synonym pairs such as *shot* and *drive*, *Direktabnahme* ('direct connection') and *Volley* ('volley'), and *tir* ('shot') and *frappe* ('shot') consist of morphologically unrelated LUs.

2. Hyponymy. A *thunderbolt* is a special kind of *shot* – specifically, a very powerful one. The same hyponymy relation holds between the German LUs *Hammer* ('hammer') and *Schuss* ('shot') and the French LUs *boulet de canon* ('cannon ball') and *tir* ('shot'). Of course, if a given LU is a hypernym of another, the relation can be extended to all synonyms of both items. In that sense, the synonym set {*Kopfball*; *Kopfstoß*} can be called a hypernym set of {*Flugkopfball*; *Kopfballtorpedo*}.

3. Translation equivalence. The German LU *Volley* and the French LU *vollée* are both translation equivalents of the English LU *volley*. As with synonymy within one language, translation equivalence across languages can, but need not be, reflected in morphological commonalities between items. An example of morphologically unrelated translation equivalents in the Shot frame is the set {*bicycle kick* / *Fallrückzieher* / *retourné*}.[12] Again, the translation equivalence relation can be extended to all members of a pair of synonym sets. For example, since *Kopfball* is a synonym of *Kopfstoß*, and *header* is a translation equivalent of *Kopfball*, *header* must also be a translation equivalent of *Kopfstoß*.

Two further types of semantic relations can be found with verbal and nominal LUs, respectively, in other parts of the vocabulary:[13]

4. Troponymy. The verbal equivalent of the hyponymy/hypernymy relation is troponymy, holding between verbs X and Y if "to X is to Y in some way" (cf. Fellbaum 1990: 285ff). This relation is also widely encountered in football vocabulary. Thus *thrash* and *beat* – both members of the Victory frame in the **Match** scene – are related to another via troponymy, because to thrash an opponent is to beat them in a very clear manner:

---

12. In this and the following synsets, English words come first, followed by German and French words. Words of the same language are separated by a semicolon, words from different languages by a slash.

13. Other semantic relations – in particular antonymy relations between adjectival LUs – have not yet been taken into account in the Kicktionary.

(17) a. [Olympique Lyonnais]$_{\text{WINNER}}$ *beat* [Fenerbahçe SK]$_{\text{LOSER}}$
[3-1]$_{\text{FINAL\_SCORE}}$ [in Istanbul]$_{\text{MATCH\_LOCATION}}$.

b. [NK Dinamo Zagreb]$_{\text{WINNER}}$ *thrashed* [Beveren]$_{\text{LOSER}}$
[6-1]$_{\text{FINAL\_SCORE}}$.

Similar relations hold, for instance, between the German verbs *ausspiel-en* ('*out-play*') and *austanzen* ('out-dance') in the Beat frame, or between the French verbs *perdre* ('lose') and *s'effondrer* ('break down') in the Defeat frame.

5. Meronymy. Nominal LUs may also be related to one another via a part/whole relationship – if X is a constituent part or a member of Y, X is a meronym of Y, and Y a holonym of X. The meronymy/holonymy relation is especially prominent in the more static scenes. Thus, many LUs belonging to frames in the **Field** scene are connected to one another via this semantic relation: the *six metre box* is a part of the *penalty box* which, in turn, is a part of the *field*; the *goalpost* is a part of the *goal*, etc. Likewise, the frames in the **Actors** scene contain many meronym/holonym pairs like English *forward – attack*, French *défense centrale* ('central defence') – *défense* ('defence') or German *Schiedsrichter* ('referee') – *Schiedsrichtergespann* ('referee team').

The question is how to supplement a scenes-and-frames hierarchy with the types of semantic relations above. One possible approach would be to extend or refine the concept of scenes and frames such that different semantic relations between LUs can be derived from their assignment to frames and/or from different relations of frames to one another or to the corresponding scenes. For example, frames could be constructed such that all the LUs in any single one of them are synonymous, and additional similarities between lexical units are represented by an appropriate relation between such minimal frames. Thus, there could be a frame Volley containing only the noun *volley*, its verbal counterpart *volley* and its German and French equivalents, another frame Header containing the noun *header*, the verb *head* etc. and a Frame Shot containing LUs like *shot*, *shoot, drive*, etc.; the Volley and Header frames could be connected to the Shot frame via a relation stating that the former are more specific versions of the latter. Up to a certain degree, this kind of solution is pursued by the Berkeley FrameNet project where the notion of 'frame inheritance' is, at least partly, related to the notion of troponymy/hyponymy between lexical units (see Ruppenhofer et al. 2006: §6).

For the Kicktionary, I decided to model these semantic relations *independently* of the scenes-and-frames structure of the resource, because I wanted to avoid having to add a further semantic dimension to existing frame and scene descriptions. Thus, I first partitioned the complete list of lexical units into synsets. The notion of a synset is borrowed from Word-Net, where it is defined as "[a] synonym set; a set of words that are interchangeable in some context" (cf. WordNet Glossary). To capture similarities in the three languages, I extended the notion of synset to include translation equivalence across languages as well as synonymy within one language.[14]

On the basis of the partition of LUs into multilingual synsets, I then established additional semantic relations between synsets, leading to three different kinds of synset hierarchies. The first is the hyponymy/hypernymy relation between nominal synsets, which yielded, for example, a taxonomic tree of multilingual terms for players' positions:[15]

(18)   {*player | Spieler | joueur*}
        {*goalkeeper; custodian | Torhüter; Torwart | gardien*}
        {*defender | Verteidiger; Abwehrspieler | arrière; défenseur*}
                {*central defender | Innenverteidiger | défenseur central*}
                        {*sweeper | Abräumer |*}
                        {*| Libero | libero*}   [. . .]

As mentioned above, the meronymy/holonymy relation is especially important for structuring lexical units in the static scenes, like those describing the playing field and its components:

(19)   {*field; pitch | Platz; Spielfeld | champ; terrain*}
        {*half | Hälfte; Spielhälfte | moitié de terrain*}
            {*penalty box; area | Sechzehner | surface de réparation*}   [. . .]
        {*touchline | Außenlinie; Seitenlinie | ligne de touche*}   [. . .]

Concerning the troponymy relation between verbal synsets, Fellbaum's (1990: 287) observation that the resulting "verb hierarchies tend to have a

---

14. This approach differs from Euro WordNet (Vossen et al. 1997), which also proposes to link synsets across different languages, but which uses an unstructured interlingual index as a separate structural entity.

15. In this tree, LUs in consecutive lines are in a hyponymy relation to one another. Thus, a *sweeper* is a (kind of) *central defender*, a *central defender* is a (kind of) *defender*, a *defender* is a (kind of) *player* and so forth.

more shallow, bushy structure than nouns" was confirmed.[16] The following is an example of such a shallow hierarchy:

(20)   {*beat*; *defeat* / *bezwingen*; *schlagen* / *battre*; *vaincre*}
          {*thrash* / *deklassieren*; *überrollen* / *écraser*; *balayer*}

7.   The Kicktionary

The Kicktionary is the result of the workflow described in the previous sections. As Table 1 shows, it currently contains close to 2,000 LUs in English, German and French:

Table 1. LUs in the Kicktionary

	English	*German*	*French*	*All*
Lexical Units (total)	599	792	535	1926
Nouns	318	451	290	1059
Verbs	248	305	201	754
Other	33	36	44	113

For each of these LUs, between one and fifteen example sentences are annotated, as Table 2 illustrates:

Table 2. Examples and annotations in the Kicktionary

	English	*German*	*French*	*All*
Examples	2374	3551	2239	8164
Examples/LU	3.96	4.48	4.19	4.24
Annotated FEs	3882	5731	3647	13260
Annotated supports	293	554	340	1187

---

16. It also seems that, in general, the problematic cases of deciding on lexical relations between LUs (including synonymy) were far more frequent in the verbal than in the nominal domain.
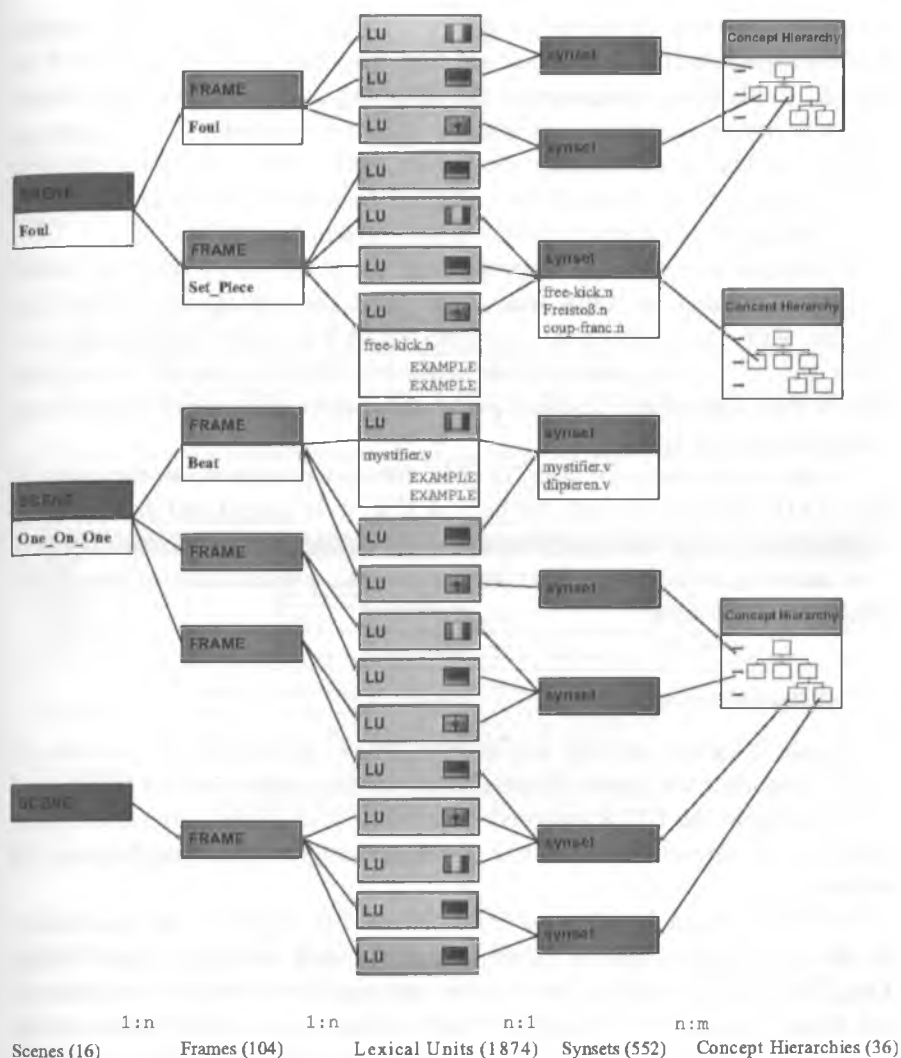
Figure 1. Organization of the Kicktionary

The basic unit of the Kicktionary is the LU, together with a set of annotated example sentences. As described above and illustrated in Figure 1 below, the list of LUs is further structured along two lines: (1) each LU is assigned to one of 104 frames, where each of these frames belongs to one of 16 scenes; (2) the list of LUs is partitioned into 552 synsets, and these synsets are further organized into a number of concept hierarchies

using the semantic relations of hyponymy/hypernymy (20 hierarchies), meronymy/holonymy (6 hierarchies) and troponymy (10 hierarchies). In contrast to all other assignments, the mapping of synsets to concept hierarchies is neither complete nor unique – i.e., whereas each LU belongs to exactly one frame and exactly one synset, and each frame to exactly one scene, some synsets may not be assigned to a concept hierarchy at all, while others may be part of two or more concept hierarchies.

For purposes of editing and processing, the Kicktionary data are stored in a small number of XML files – one large file containing all the LUs together with their annotated examples as well as their assignments to a frame and to a synset, one file containing the different concept hierarchies, and 16 files containing descriptions of the scenes and information about what frames they consist of.

For presentation to the user, HTML files are generated on the basis of these XML files (mostly with the help of XSL style sheets) and disseminated via the freely available Kicktionary website (http://www.kicktionary.de). The following subsections describe the HTML presentation of the Kicktionary in more detail.

7.1.   Presentation of LUs

As Figure 2 shows, the top line of each entry indicates the base form of the LU together with part of speech information and to which frame and which scenario the LU is assigned. The frame and scene names are hyperlinked to the presentations of the corresponding entities (see Section 7.2 below).

This description is followed by a list of FEs used in the annotation of the LU. Apart from a label indicating their semantic type[17] (e.g., 'On_The_Field_Location'), no further information about FEs is given at this level – since FEs are defined with respect to a superordinate scene, and not to individual LUs, I decided that the level of scenes is the best place to provide this definition (see next section).

The annotated example sentences are displayed in the center of the screen. Annotated FEs are indicated by a set of square brackets, with the FE name appended as a subscript. The form of the LU is printed in bold,

---

17. This assignment of FEs to semantic types – a kind of broader ontological classification of FEs (see Schmidt 2006) – is a further level of structure in the resource which was, however, not fully developed, and is, therefore, not treated in this paper.

drill.v
■   Scenario Shot Frame Shot

BALL [Ball]
PATH [On_The_Field_Location]
SHOOTER [Player]
SOURCE [On_The_Field_Location]
TARGET [On_The_Field_Location]

---

1. [Henry]$_{SHOOTER}$ cut inside from the right and laid the ball in to Ljungberg inside the penalty area, collecting the Swede's back-heeled return to **drill** [into the corner of the net]$_{TARGET}$. [1077218 / p6]

2. Poland's seventh came five minutes before the end when Tomasz Rzasa again exploited the visitors weak right side and cut back to substitute [Marek Saganowski]$_{SHOOTER}$ who spun in the area before **drilling** [the ball]$_{BALL}$ [home]$_{TARGET}$. [75339 / p8]

3. In the final minutes of the opening period Simão, Petit and Nuno Assis all took turns to try to level the score with audacious efforts from distance while CSKA's only attempt of note during this period fell to [Rolan Gusev]$_{SHOOTER}$, who also **drilled** [wide]$_{TARGET}$ [from long range] $_{SOURCE}$. [79954 / p6]

4. Krzynówek deservedly scored himself on 72 minutes, when [he]$_{SHOOTER}$ came off the left wing and **drilled** [high]$_{PATH}$ [past Kramarenko] $_{TARGET}$ [from the angle of the area]$_{SOURCE}$. [75339 / p7]

drill			Henry		into the corner of the ne...
drilling	the ball		Marek Saganowski		home
drilled			Rolan Gusev	from long range	wide
drilled		high	he	from the angle of the are...	past Kramarenko

Synonyms	hämmern.v / jagen.v / knallen.v / dreschen.v
	blast.v / fire.v / rifle.v / drill.v
	catapulter.v / propulser.v
Troponyms [Move_Ball]	draufhalten.v / abziehen.v / schießen.v
	kick.v / shoot.v
	shooter.v / tirer.v / frapper.v

Figure 2. Presentation of the LU *drill*

and supports are underlined. Following each example sentence, information is given about the corpus text from which it was excerpted. Clicking on this information will take the user to a full text presentation of the match report in question.

A second, schematic representation of the examples in the form of a table allows users to study commonalities and differences between examples with respect to the surface forms of LUs and their FEs. The table hides all but LUs and FEs and lists the FEs name-by-name instead of in order of appearance in the sentence.

The lower part of the screen shows information about semantic relations of a LU with other LUs in the Kicktionary. First, the corresponding synset is displayed, providing the user with hyperlinks to all existing synonyms in the same language and translation equivalents in the other

languages. Where appropriate, this is followed by a similar display of superordinate synsets from one or more of the concept hierarchies. Additionally, users are given a link to a complete presentation of the respective concept hierarchy (see below) and can explore hyponyms, co-hyponyms, meronyms and troponyms via this level.

7.2.   Presentation of scenes and frames

Recall that in the Kicktionary, several frames make up a scene. When representing this relation, it is important to keep in mind that a scene, by definition, corresponds to a kind of knowledge that is not (or not exclusively) linguistic in nature. From the point of view of a dictionary, this means that a textual description, a short film or a schematic diagram may all be equally adequate representations of a scene. In fact, if the role of a scene as an interlingual mediator in the organization of a multilingual vocabulary is emphasized, there are even good reasons to prefer non-linguistic forms of presenting a scene over linguistic ones.

In its present form, the Kicktionary illustrates most scenes with one or more schematic diagrams such as the **Shot** scene in Figure 3:
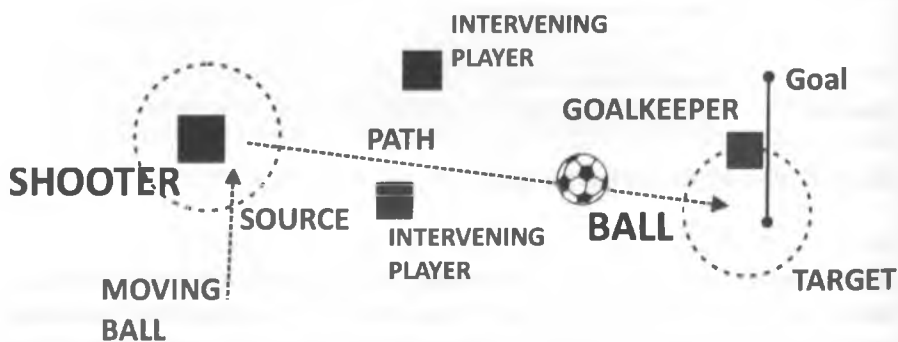
Figure 3. A schematic diagram of the **Shot** scene

The diagram in Figure 3 shows the main actors of the **Shot** scene (and the corresponding FE names), and represents their spatial constellation on the field while conveying a general idea of the temporal dynamics of the scene. A short film, possibly with appropriate subtitles and/or some graphical means of highlighting certain portions, would probably serve the same purpose in an even better way. In some instances, I also found that a scene or a part of a scene can be very adequately illustrated by a single photo or drawing which captures in some way a prototypical mental

image associated with that scene. This was the case, for instance, for the `Celebration` frame in the **Goal** scene and for the **Substitution** scene as in Figure 4:

Figure 4. Images illustrating the `Celebration` frame and **Substitution**[18] scene, respectively

The graphic information is supplemented with a prose description of the scene, which lists the FEs, explains their roles in the action, and sketches the typical course of events in the scene. After the scene is explained in that way, the user is given links to the various corresponding frames, as is shown in Figure 5.

The **Shot** scene is centered around the event of a player directing the ball to a target on the field. Typically, the target is the opponent's goal, and the shot is carried out with the intention of scoring a goal. The main protagonist of the scene is the SHOOTER. Using a PART OF HIS BODY, the shooter directs the BALL towards the opponent's goal. The ball moves from the SOURCE location on the field along a PATH to a TARGET location. In some cases, the MOVING BALL (typically a pass from a team-mate) that brought the shooter into a position to carry out the shot can be mentioned. Sometimes, a shot is construed as the final stage of a MOVE by the shooter's team.

The frame `Shot` contains LUs which describe a shot from the shooter's point of view. The `Finish` frame contains LUs that construe a shot as the last stage of a move by the shooter's team. [...]

Figure 5. The text introducing the **Shot** scene

---

18. Images taken from [http://www.drblank.com/slaw3.htm].

Flick_On [Scenario: Pass]

1. RECIPIENT
2. PASS
3. SECOND_RECIPIENT
4. BALL
5. PART_OF_BODY
6. DIRECTION
7. FLICK_ON_TARGET

*	1	2	3	4	5	6	7
verlängern.v	■	■	■				■
Verlängerung.n	■	■	■	■	■		
weiterleiten.v	■	■	■				
flick_on.n		■	■	■			
flick_on.v	■	■					
touch_on.v	■	■					
prolonger.v	■	■	■	■			

Figure 6. Schematic overview of the content of the frame `Flick_On`

Given that all the contextual knowledge needed to understand the definition of a certain frame is already provided at the level of the superordinate scene, the presentation of a frame is restricted to a schematic overview of the relevant LUs and the FEs encountered with them. In Figure 6, this is done in the form of a table in which the LUs of a frame (sorted first by language, then alphabetically) are listed row-by-row and the FEs used in the annotation are listed column-by-column. The table cells indicate which FE is encountered with which LU. Clicking on any of the LUs will take the user to the corresponding LU representation.

7.3. Other elements of the presentation

In addition to the information outlined above, the web version of the Kicktionary provides a separate visualization of the organization of LUs into hierarchies of synsets (similar to WordNet, see Fellbaum 1998). There is a two-way-link between these representations and the representations of individual LUs so that a user can navigate from a given LU to one of its hyponyms or co-hyponyms via such a hierarchy, as illustrated in Figure 7.

The Kicktionary also provides a full-text display of the corpus texts, which can be accessed via the link provided in the example section of the LU presentation (see Figure 2 above). This allows users to study the larger

Individual_Actors [Hypernymy]

Individual_Actor

Akteur.n   Spieler.n
player.n
joueur.n

Keeper.n   Schlussmann.n   Torhüter.n   Torwart.n
custodian.n   goalkeeper.n   keeper.n
gardien_de_but.n   gardien.n   portier.n

Feldspieler.n

Abwehrspieler.n   Verteidiger.n
defender.n
arrière.n   défenseur.n

Innenverteidiger.n
central_defender.n   centre-back.n   full-back.n
défenseur_central.n

Abräumer.n
sweeper.n

Figure 7. Presentation of the 'Individual_Actors' concept hierarchy

context in which the annotated example sentences appear. Finally, several means for top-level navigation provide the user with points for exploring the full list of LUs and their various forms of organization. For a bottom-up access to the Kicktionary, a simple alphabetical list of LUs, separated by language, is provided. Alternatively, users can start with an annotated parallel text in which occurrences of LUs are linked to the respective entries in the resource, as is shown in Figure 8.

For top-down access, the user can either start with an overview of scenes and frames or with a list of concept hierarchies, as Figure 9 illustrates.

English text	German text	French text
Barcelona glee in Glasgow	**Barcelona mit historischem Sieg**	**Le Barça jubile à Glasgow**
Tuesday , 14 September 2004	Dienstag , 14. September 2004	Mardi , 14 septembre 2004
By Alex O ' Henley at Celtic Park	Von Alex O ' Henley aus dem Celtic Park	Par Alex O ' Henley à Celtic Park

FC Barcelona became the first visiting team to win a UEFA Champions League match at Celtic FC as goals from Deco, Ludovic Giuly and the homecoming Henrik Larsson secured maximum points in their Group F opener .

Der FC Barcelona hat als erste Mannschaft ein UEFA Champions League-Spiel bei Celtic FC gewonnen. Beim 3:1-Erfolg im Celtic Park trafen Deco, Ludovic Giuly und "Heimkehrer" Henrik Larsson und sicherten so den Katalanen zum Auftakt in der Gruppe F drei Punkte.

Le FC Barcelona est la première équipe à remporter un match d'UEFA Champions League à Celtic Park. Deco, Ludovic Giuly et Henrik Larsson, qui effectuait son grand retour, ont marqué et offrent les trois points au Barça dans le Groupe F.

Larsson clincher

Entscheidung durch Larsson

Larsson buteur

Larsson's goal, on his return to the club where he scored 242 goals in a seven-year spell, sealed a victory which had looked in doubt after Chris Sutton grabbed a dramatic equaliser for Celtic just short of the hour mark . Barcelona coach Frank Rijkaard had named Larsson as a substitute with Brazilian ace Ronaldinho coming in to form a front three with Ludovic Giuly and Samuel Eto ' o.

Larssons Tor gegen seinen ehemaligen Verein, für den er in sieben Jahren 242 Mal getroffen hatte, sorgte für die endgültige Entscheidung. Zwischendurch nahm die Partie dramatische Züge an, als nach etwa einer Stunde Chris Sutton der umjubelte Ausgleich gelang. Barcelonas Trainer Frank Rijkaard hatte Larsson zunächst nur auf die Ersatzbank gesetzt, um im Angriff mit dem brasilianischen Star Ronaldinho sowie Ludovic Giuly und Samuel Eto'o zu beginnen.

Le but de Larsson , pour son grand retour au club pour lequel il a marqué 242 buts en sept ans , scellait une victoire qui ne semblait pas acquise après l'égalisation de Chris Sutton à l'approche de l'heure de jeu . Le coach du Barça , Frank Rijkaard, décidait de reléguer Larsson sur le banc des remplaçants, alors que le prodige brésilien Ronaldinho faisait son entrée pour former un trio de tête avec Ludovic Giuly et Samuel Eto'o.

Figure 8. An annotated parallel text, linked to the lexical resource

Chance	Chance (10 LUs)	**Hypernymy / Hyponymy**
3 frames 26 lexical units	Create_Chance (7 LUs) Miss_Chance (9 LUs)	Individual_Actors (143 lexical units) Team_Parts (20 lexical units)
Motion	Ball_Bounce (13 LUs) Ball_Escape (3 LUs)	Moving_Balls (175 lexical units) Matches (38 lexical units)
7 frames 88 lexical units	Ball_Land (12 LUs) Ball_Move (19 LUs) Goalkeeper_Advance (6 LUs) Player_Move (10 LUs) Player_Move_With_Ball (25 LUs)	Goals (37 lexical units) One_On_One_Situations (16 lexical units) Match_Outcomes (27 lexical units) Offences (43 lexical units)
Substitution	Bring_Off (8 LUs) Bring_On (12 LUs)	Individual_Actions (76 lexical units) Substitutions (9 lexical units)
3 frames 27 lexical units	Substitute (7 LUs)	Dead_Ball_Situations (48 lexical units)

Figure 9. Overview of scenes, frames and concept hierarchies

8. Evaluation

Since the Kicktionary can, in essence, be regarded as a multilingual, domain-specific adaptation of the methodology underlying the FrameNet project (Fillmore et al. 2003), a large part of the discussion in this section is concerned with a comparison of these two resources.

8.1. The multilingual aspect

Concerning the *construction* of a multilingual resource, the strategy of carrying out a scenes-and-frames analysis on several languages simultaneously has proven feasible, generally supporting Boas' (2005a) claim that semantic frames are useful as interlingual representations. Concerning the *use* of the Kicktionary for translation or similar tasks, examples like the ones discussed in Section 4 provide further evidence that diverse cases of cross-linguistic (non-)correspondences can be partly accounted for in frame semantic terms in a way that should be transparent and beneficial to dictionary users.

Furthermore, the concept of a scene provides a theoretically substantiated justification for introducing non-linguistic methods of description into dictionaries. As has been argued in the lexicographic literature (e.g. Storrer 2001), and as existing commercial electronic dictionaries show, the fact that computer technology facilitates the use of pictures, diagrams, films etc., alongside textual material opens interesting perspectives for monolingual as well as for multilingual dictionaries. Because Frame Semantics is, among other things, concerned with systematically relating linguistic forms to non-linguistic knowledge, a scenes-and-frames analysis can help define what kinds of information such multi-medial elements should convey, and determine at which level a resource should place it.

8.2. The domain-specific aspect

To my knowledge, the Kicktionary is one of the first attempts to apply frame semantic principles systematically to the vocabulary of a specific domain. This has a number of advantages.

First, football is a particularly rewarding domain because most of its scenes can be associated in a straightforward manner with concrete mental images – the notion of a scene (as understood here) is arguably much more intuitively applicable for LUs like *foul*, *goal* and *scissors kick* than it is for many parts of the general vocabulary which denote more abstract concepts, such as *depend*, *necessity* or *tolerant* (all from the FrameNet

database). For similar reasons, difficulties in distinguishing literal and metaphorical uses of words hardly arise in the language of football.

Second, restricting the analysis to a specific domain also entails a limitation to a closed set of LUs, which means that there is a definable line beyond which LUs will not be taken into account because they fall outside the domain.[19] This limitation can be seen as an advantage from a methodological point of view: it allows for a manner of proceeding in which first a reasonably extensive (if not complete) list of LUs and example sentences is extracted from the corpus. Scenes and frames are then built on top of that list and the completeness of the resulting structure is continually checked with respect to the list.[20] This is different from FrameNet, which proceeds frame by frame, selecting candidate LUs for frames mainly through linguistic introspection, and only then consulting the corpus for evidence in favor of the tentative analysis.[21] An advantage of the Kicktionary methodology is that it makes it much easier to estimate the effects of an individual decision on the resource as a whole. For instance, many of the problems discussed in Section 5 were resolved[22] by considering which one of a number of potential alternative analyses would result in a more economic,

---

19. In the case of the Kicktionary, the set of lexical units was further limited by the relatively small size of the corpus – between 250,000 and 1,000,000 words for each language as compared to the 100,000,000 words of the BNC on which the FrameNet database is based. With few exceptions, words that could not be found in this small corpus were not considered for integration into the resource.

20. This is of course a simplified picture. In reality, the list could only be assembled with the help of a preliminary scenes-and-frames analysis of the football domain, which was then "thrown away" and rebuilt from scratch. The crucial point, however, is that developing scenes and frames and determining the LUs which are to become part of them can be regarded as two separate processes for the Kicktionary whereas they are inseparably interwoven for FrameNet.

21. In a discussion on the lexicography mailing list, this methodology is criticized as follows: "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/]

22. And, conversely, some of these problems arose exactly *because* the scenes-and-frames structure of the Kicktionary was constructed to accommodate the entirety of LUs found in the corpus. Proceeding frame-by-frame always involves a certain risk of leaving exactly those LUs unanalysed that are ambivalent with respect to their framing characteristics.

homogeneous, balanced or useful overall structure of scenes and frames; and, of course, such a process presupposes that the majority of the LUs to be integrated into the structure be known at the time of analysis.

8.3. Scenes and frames, frame inheritance, and other entities and concepts

Although both resources are constructed on the basis of frame semantic principles, the Kicktionary and FrameNet differ in important points both with respect to their form, i.e., the actual data structures they use to represent their respective frame semantic analyses, and with respect to their content.

For example, FrameNet takes a much more comprehensive approach to the annotation of examples. Each LU is illustrated with a much larger number of sentences from the corpus than in the Kicktionary, and the annotation of these sentences is also much more extensive: in addition to the information about FEs, their grammatical functions (e.g., object, dependent) and their phrase types (e.g., noun phrase, prepositional phrase) are recorded. Time restrictions precluded this level of detail for the Kicktionary. Similarly, FrameNet uses the concept of null instantiation of FEs for "FEs that are conceptually salient, [but] do not show up as lexical or phrasal material in the sentence for annotation" (Ruppenhofer et al. 2006: §3.2.3). The Kicktionary does not make use of null instantiation; this does not mean that it was considered unimportant, but only that I lacked the time to integrate it into my analyses. The same holds for a number of other details of the FrameNet database like the notion of coreness, the bundling of FEs into core-sets or the annotation of extra-thematic FEs (see Fillmore et al. 2003).

Another difference between the two resources is that in FrameNet, the only top-level structural entities are frames (including specific types such as non-lexical frames, non-perspectivized frames, see Ruppenhofer et al. 2006: §6.2), which are related to one another via an elaborate system of frame-to-frame relations (e.g., inheritance, causative_of, inchoative_of, subframe, etc.). In contrast, the scene is the Kicktionary's top level entity, and it is explicitly understood as a unit substantially different from (and superordinate to) that of a frame. Each frame is associated with exactly one such scene, and this frame-to-scene assignment is also the only explicit way of relating frames to one another. Whereas a similar relationship can be expressed in FrameNet by connecting a lexical frame to a non-lexical frame via the "subframe" relation, nothing in the design of the FrameNet

database *requires* such a frame-to-scene-assignment.[23] The notion of a scene and the distinction between scenes and frames are thus much more central to the Kicktionary than they are to FrameNet.

8.4.   Frame Semantics and other analyses

Work on the Kicktionary suggests that an ideal lexicographic analysis for the purpose of dictionary-making will require both a methodologically motivated restriction of the role of Frame Semantics to certain areas of the vocabulary and an appropriate use of other approaches to semantic analysis.[24] By organizing the vocabulary of football language both in a scenes-and-frames hierarchy and in a WordNet-like system of synsets and concept hierarchies, the Kicktionary has partly explored the second of these requirements. One observation in this respect is that WordNet-style analyses often seem to be most profitable in precisely those areas where frame semantic analyses are less intuitively applicable or less informative (see also Boas 2005b). For instance, I argued in Section 5 that a scenes-and-frames analysis of LUs referring to parts of the playing field is made difficult by the fact that the notion of perspective is not easily applied to such a static "scene". At the same time, example (19) shows that this set of LUs can be very intuitively structured on the basis of semantic relations like synonymy and meronymy. Conversely, it was found that troponymy between verbal LUs seems to be a semantic relation that is more difficult to detect or analyze and/or less widely encountered than hyponymy or meronymy relations between nominal LUs. In this area, then, the kind of relation that a scenes-and-frames analysis establishes between verbal LUs may be the more useful one from the point of view of a dictionary user. Since real conflicts between the two approaches were not encountered, a tentative conclusion to be drawn from these observations is that Frame-Net- and WordNet-style analyses should be viewed more as complementary, rather than in opposition to each other.[25]

---

23. And, in fact, most lexical frames in the FrameNet database are *not* related to a superordinate non-lexical frame.
24. As Fillmore (1978) states for semantic theory in general: "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."
25. That is, there were no cases where an analysis according to one approach would positively contradict or be incommensurable with an analysis according to the other approach.

9. Summary and outlook

In this paper I discussed the theoretical background and the workflow underlying the Kicktionary, a multilingual, domain-specific lexical resource based on Frame Semantics. My comparison of the structure and content of the Kicktionary with more general lexical resources such as FrameNet and WordNet has resulted in several insights. First, a hierarchy of scenes and frames is an efficient way of grouping sense-related *domain-specific* vocabulary items on a level which abstracts over linguistic form, and thus constitutes a connection between linguistic and "world" knowledge. Second, FrameNet-style annotations provide an efficient way of including empirical language material in electronic dictionaries. Systematically relating the labels used in these annotations to the hierarchy of scenes and frames opens further possibilities for the dictionary user to discover and exploit relationships between lexical items. Third, the scenes-and-frames approach lends itself very well to the construction of a *multilingual* resource that can be helpful in various translation tasks. Fourth, decisions about frame and scene membership of a LU are not always straightforward. Often, pragmatic considerations about the economy of the dictionary design are a way of dealing with such difficulties. Fifth, a scenes-and-frames analysis is easier and more fruitful in those areas of the vocabulary which deal with dynamic activities than in more static areas. For the latter, WordNet-style concept hierarchies seem like the more intuitive and more useful approach. As such, a scenes-and-frames analysis and a WordNet style analysis of the lexicon are complementary to each other. Finally, the concept of a scene providing information about prototypical events gives dictionary writers a useful place for integrating multi-media elements like pictures or films that aid in the comprehension of words in foreign languages.

When constructing multilingual lexical resources, it is important to keep in mind that football is probably not a prototypical case of a special domain. Other specialized domains are likely to exhibit larger, more deeply nested, and more systematic taxonomic systems. Dynamic aspects, and hence the benefits of a scenes-and-frames organization of the lexicon, may play a less prominent role in their analysis. In contrast to football language, they will tend to avoid, rather than abound with, synonymy and near-synonymy so that the task of establishing links between lexical items is different. Work by Dolbey et al. (2006) on "Bio FrameNet" is an example of such a more typical specialized lexical resource.

At this point in time, the Kicktionary is complete in the sense that a reasonably large number of LUs from the football domain has been ana-

lyzed and integrated into the described architecture.[26] It is also complete in the sense that this architecture is accessible via a website. There are, however, various ways in which it could be improved and extended.

First, an extension of the corpus is likely to uncover new LUs and a larger corpus could be used to increase the number of annotated examples for existing LUs. In both cases, the additional material may make it necessary to remodel parts of the scenes-and-frames hierarchy and parts of the concept hierarchies. Further text materials from the UEFA website (about 250,000 tokens for English, French and German) have been acquired for this purpose and are presently being processed.

Second, user feedback for the Kicktionary website should make it possible to evaluate the quality of the resource and its presentation. One possible way of improving the presentation might be the inclusion of additional films and pictures into the description of scenes.

Third, the existing architecture, together with the concordancing and annotation tool developed for the analysis, should make it relatively easy to supplement the Kicktionary with data from other languages. Italian, Portuguese, Spanish, Russian and Japanese corpus materials are available for lexicographers interested in producing versions for these languages.

Finally, I would like to suggest that the Kicktionary should be regarded as a promising test case for the development and application of methods for collaborative creation of specialized multilingual lexical resources, because (1) football is a well-delimited special domain with a large, but manageably-sized vocabulary, and (2) contrary to many other specialized areas, it is not too difficult to find "experts" who are competent users of that vocabulary (in different languages) and who may be able and willing to contribute to such a collaborative effort either as lexicographers or as evaluators of the resulting resource.[27] First steps towards a client-server architecture in which dictionary creators and dictionary users can work together to construct an improved version of the Kicktionary have already been taken.

---

26. "Reasonably large" means that (a) the number of lexical units in the Kicktionary is considerably higher than in comparable printed dictionaries (e.g. Yıldırım 2006, Colombo et al. 2006) and that (b) a further analysis of the corpus would turn up no or very few additional LUs.

27. So far, online feedback shows that the Kicktionary seems indeed capable of getting both linguists and laymen interested in lexicography.

References

Boas, Hans C.
  2005a   Semantic frames as interlingual representations for multilingual
               lexical databases. In: *International Journal of Lexicography*
               18.4: 445–478.
Boas, Hans C.
  2005b   From theory to practice: Frame Semantics and the design of
               FrameNet. In: S. Langer and D. Schnorbusch (eds.), *Semantik
               im Lexikon*, 129–160. Tübingen: Narr.
Colombo, Roberta, Klaus Heimeroth, Olivier Humbert, Michael Jackson, Frank
               Kohl, and Josep Ràfols
  2006   *PONS Fußballwörterbuch*. Stuttgart: Ernst Klett Verlag.
Dolbey, Andrew, Michael Ellsworth, and Jan Scheffczyk
  2006   BioFrameNet: A domain-specific FrameNet Extension with
               links to biomedical ontologies. In: *Proceedings of the Interna-
               tional Workshop "Biomedical Ontology in Action"*, November 8,
               2006, in Baltimore, MD.
Fellbaum, Christiane
  1990   English verbs as a semantic net. In: G.A. Miller et al. (eds.),
               WordNet – an Online Lexical Database. *International Journal
               of Lexicography* 3.4: 278–301.
Fellbaum, Christiane
  1998   *WordNet: an electronic lexical database*. Cambridge: MIT Press.
Fillmore, Charles J.
  1977a   The case for case reopened. In: P. Cole and J. Sadock (eds.),
               Syntax and Semantics 8: Grammatical Relations, 59–82. New
               York: Academic Press.
Fillmore, Charles J.
  1977b   Scenes-and-frames semantics, linguistic structures processing. In:
               A. Zampolli (ed.), *Fundamental Studies in Computer Science*,
               No. 59, 55–88. Dordrecht: North Holland Publishing.
Fillmore, Charles J.
  1977c   Topics in lexical semantics. In: R. Cole (ed.), *Current Issues in Lin-
               guistic Theory*, 76–138. Bloomington: Indiana University Press.
Fillmore, Charles J.
  1978   On the organization of semantic information in the lexicon. In:
               D. Farkas et al. (eds.), *Papers from the Parasession on the Lexi-
               con*, Chicago Linguistic Society, April 14–15, 1978. Reprint in:
               Fillmore, Charles J., *Form and Meaning in Language: Volume I,
               Papers on Semantic Roles*, Stanford: CSLI Publications, 261–
               289.
Fillmore, Charles J., Christopher Johnson, and Miriam R.L. Petruck
  2003   Background to FrameNet. In: *International Journal of Lexico-
               graphy* 16.3: 235–250.

Gross, Gaston
  2002   Comment décrire une langue de spécialité? In: *Cahiers de lexico-
                logie: revue internationale de lexicologie et lexicographie* 80: 179–
                200.
Petruck, Miriam R.L.
  1996   Frame Semantics. In: J. Verschueren et al. (eds.), *Handbook of
                Pragmatics*, 1–13. Amsterdam/Philadelphia: John Benjamins.
Ruppenhofer, Josef, Michael Ellsworth, Miriam R.L. Petruck, and Chris Johnson
  2006   FrameNet: Theory and Practice. http://framenet.icsi.berkeley.
                edu/book/book.html
Schmidt, Thomas
  2006   Interfacing lexical and ontological information in a multilingual
                soccer FrameNet. In: *Proceedings of OntoLex 2006 – Interfacing
                Ontologies and Lexical Resources for Semantic Web Technolo-
                gies*. Genoa, Italy, May, 24–26, 2006.
Seelbach, Dieter
  2001   Das kleine multilinguale Fußball-Lexikon. In: W. Bisang and
                G. Schmidt (eds.), *Philologica et Linguistica. Historia, Pluralitas,
                Universitas*, 323–350. Trier.
Seelbach, Dieter
  2002   La traduction des verbes avec adverbes appropriés et des verbes
                à particule allemands. In: *Traduire au XXIeme siecle: Tendances
                et perspectives, Proceedings 2002*, 504–515. Facultés des lettres
                UATH Athens.
Seelbach, Dieter
  2003   Separable Partikelverben und Verben mit typischen Adverbialen.
                Systematische Kontraste Deutsch-Französisch / Französisch-
                Deutsch. In: U. Seewald-Heeg et al. (eds.), *Sprachwissenschaft,
                Computerlinguistik, Neue Medien*, 103–115. Königswinter.
Storrer, Angelika
  2001   Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hyper-
                textkonzepts in der Lexikographie. In: I., Lemberg, B. Schröder,
                and A. Storrer (eds.), *Chancen und Perspektiven computerge-
                stützter Lexikographie. Hypertext, Internet und SGML/XML für
                die Produktion und Publikation digitaler Wörterbücher*, 88–104.
                Tübingen: Niemeyer.
Vossen, Piek, Pedro Díez-Orzas, and Wim Peters
  1997   Multilingual design of EuroWordNet. In: P. Vossen, N. Calzo-
                lari, G. Adriaens, A. Sanfilippo, and Y. Wilks (eds.), *Proceed-
                ings of the ACL/EACL-97 workshop on Automatic Information
                Extraction and Building of Lexical Semantic Resources for NLP
                Applications*. Madrid, July 12th, 1997.
                WordNet Glossary: http://wordnet.princeton.edu/gloss
Yıldırım Kaya
  2006   *Fußballwörterbuch in 7 Sprachen*. Kauderwelsch (203). Osna-
                brück: Reise-Know-How Verlag Peter Rump GmbH.