A Representation Framework for Cross-lingual/Interlingual Lexical Semantic Correspondences

Yoshihiko Hayashi
Osaka University, Japan
hayashi@lang.osaka-u.ac.jp

Abstract

This paper proposes a framework for representing cross-lingual/interlingual lexical semantic correspondences that are expected to be recovered through a series of on-demand/on-the-fly invocations of a lexical semantic matching process. One of the central notions of the proposed framework is a *pseudo synset*, which is introduced to represent a cross-lingual/multilingual lexical concept, jointly denoted by word senses in more than one language. Another important ingredient of the proposed framework is a framework for *semantifying bilingual lexical resource entries*. This is a necessary substep when associating and representing corresponding lexical concepts in different languages by using bilingual lexical resources. Based on these devices, this paper further discusses possible extensions to the ISO standard lexical markup framework (LMF). These extensions would enable recovered correspondences to be organized as a dynamic *secondary language resource*, while keeping the existing primary language resources intact.

1 Introduction

As the world goes more global, the demand for multilingual lexical semantic resources has increased. A central approach to realize such a multilingual resource has been nicely demonstrated by the EuroWordNet (Vossen 2004) and the succeeding it, Global WordNet Grid project[1]. In these projects, the goal is to build a worldwide grid of wordnets by means of interlingual pivots. While we may assume that the grid is static and stable in its nature, *dynamic lexical resources* (Calzolari 2008) are possible, provided a variety of language resources are wrapped as Web services[2] and are accessible on a service infrastructure. For example, a virtually *combined lexicon*[3] can be evolutionarily realized by opportunistically associating semantically corresponding entries in the relevant lexical resources.

However, existing frameworks for modeling and representing lexical resources are not applicable to this new type of lexical resource in their current configurations. For example, while the ISO lexical markup framework (LMF)[4] provides useful constructs to represent a range of lexicons, it still concentrates on modeling one lexical resource at a time, and does not provide effective devices to integrate different types of lexical resources into a single combined resource. This has motivated us to develop a framework for representing cross-lingual/interlingual lexical semantic correspondences that may be recovered through a series of on-demand/on-the-fly invocations of a lexical semantic matching process that underlies combined lexicon access services.

The central concept of the framework is the notion of *pseudo synset*, which is introduced to represent a cross-lingual/multilingual lexical concept, jointly denoted by words in more than one language. As the name implies, it inherits and extends the constituting principle of wordnets: a lexical concept is

[1] http://www.globalwordnet.org/gwa/gwa_grid.htm

[2] We use the term *servicize* to mean the wrapping of a static language resource as a dynamic Web service, which provides a standardized application program interface (API).

[3] Hartmann(2005) discusses a range of *hybrid dictionaries*, which includes, for example, *monolingual cum interlingual dictionary*.

[4] Standardized as ISO 24613:2008.

defined as a set of synonymous word senses. Another component of the proposed framework is a framework for *semantifying* bilingual lexical resource entries, which is a necessary substep for associating and representing corresponding lexical concepts in different languages by using bilingual lexical resources.

This paper starts with a motivating example and a look at how to represent the abovementioned components in the example. This paper then discusses possible extensions to the ISO LMF, which would enable recovered cross-lingual/interlingual correspondences to be organized as a *dynamic* language resource. This dynamic resource is *secondary*, because it is created on top of the existing *primary* language resources. Here it should be noted that this secondary language resource can be enriched and expanded, gradually *evolving* in a collaborative Web service environment.

2 A Motivating Example and Representations

Figure 1 shows our motivating example, depicting five direct cross-lingual lexical semantic correspondences: a Japanese word *kawa* can be translated into either *river* or *stream* in English; *river* is associated with either of *rivière* or *fleuve* in French, depending on where the river flows into; *stream* is associated only with *rivière* in French.

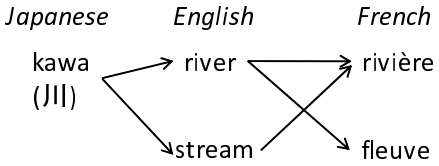

Figure 1: Motivating Example.

Situations similar to this one would be brought about, for example, by invoking a lexical access service on a Web-based linguistic service infrastructure. More specifically, think of a dictionary service that implements a virtually combined dictionary. One user of this service might like to find the meaning of the Japanese word *kawa* (by consulting a Japanese lexical semantic resource) and then want to know the equivalents in English (by consulting a bilingual dictionary); another user may want to look for French counterparts of *river*. To fulfill these requirements, a computational lexical semantic matching process behind the dictionary service should be invoked in an on-demand and on-the-fly manner, if the relevant cross-lingual semantic correspondences are unknown to it. These invocations of the matching process can induce possible indirect lexical semantic correspondences: for example, between *kawa* and *rivière*, via *river*.

2.1 Problems with a Possible LMF Representation

The LMF *NLP multilingual notation extension* (Francopoulo et al. 2009) is devised to model and represent lexical semantic correspondences across languages. We can use this device to model and represent the situation in the motivating example, as shown in Fig. 2, which makes use of the `Sense Axis` construct. Actually, this figure has been created from a figure presented in (Francopoulo et al. 2009) by adding the following: a Japanese `Sense` node associated with *kawa*; an English `Sense` node associated with *stream*; and a `Sense Axis` node that links the Japanese `Sense` node to the two English `Sense` nodes. Although this configuration seems to be natural, several questions may arise, including:

- How can we represent an indirect correspondence that could be dynamically derived or inferred from a combination of direct correspondences? For example, should the derivable indirect correspondence between *kawa* and *fleuve* also be represented by adding the `Sense Axis` and `Sense Axis Relation` constructs? Or should we introduce another `Sense Axis` node, which, as an interlingual pivot, aggregates all the corresponding senses?

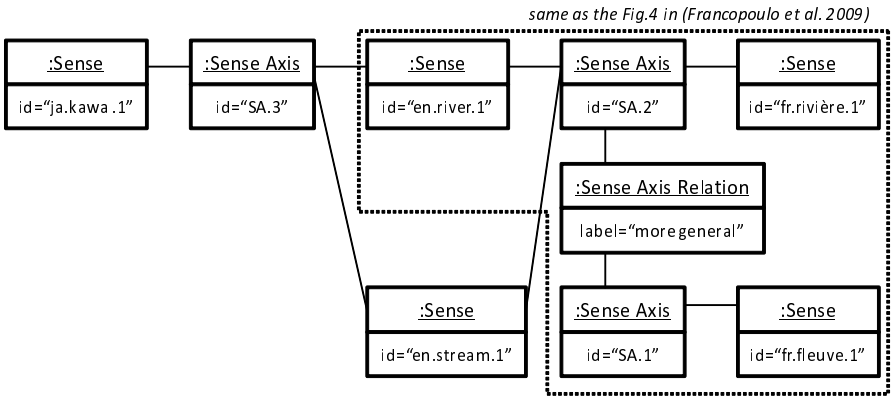

Figure 2: Straightforward LMF Representation of the Motivating Example.

- How and where should the details of a matching process be encoded? This is particularly crucial for a dynamic resource, so that the potential user is able to assess the reliability of the resource.
- Is the introduction of the `Sense Axis Relation` instance with the label "more general" necessary or adequate? The LMF specification states that a `Sense Axis Relation` instance should be introduced if the correspondence is not direct (partially equivalent). However, in our scenario, it is reasonable to expect that the lexical semantic relation between *rivière* and *fleuve* has already been encoded somewhere in an existing French lexical semantic resource. This suggests that the introduction of the `Sense Axis Relation` might be redundant.

2.2 Proposed Representation: Overview

Figure 3 shows the conceptual overview of the proposed representation for the motivating example in consideration of these questions. In this representation, we have eight nodes, each depicted by a shaded round rectangle node. Each of these nodes is classified as a *cross-lingual pseudo synset* (`CP_Synset`) node (marked by a number) or a *multilingual pseudo synset* (`MP_Synset`) node (marked by a Greek letter). While the former represents a directed cross-lingual correspondence between two senses, the latter shows a set of multilingual word senses that may share an intersectional concept across the languages. For example, the `CP_Synset` node labeled "1" represents a concept denoted by senses of *kawa* and *stream*, along with the depicted direction. The node marked α indicates a concept jointly denoted by the multilingual sense set: $\{$*kawa*, *stream*, *rivière*$\}$.

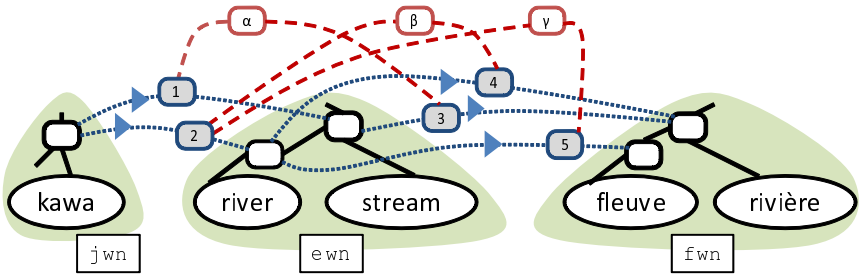

Figure 3: Conceptual Overview of the Proposed Representation for the Motivating Example.

Given the previously mentioned use case scenario, we presuppose that two types of lexical resources already exist, and that they are made accessible by appropriate Web service interfaces:

- Three WordNet-type monolingual lexical semantic resources for Japanese (`jwn`), English (`ewn`) and French (`fwn`) are assumed. We assume that they are modeled and represented using the LMF NLP semantics extension.

- Although not explicitly depicted in this figure, two bilingual lexical resources for Japanese-to-English (`j-to-e`) and English-to-French (`e-to-f`) are assumed. They are assumed to be modeled and represented by employing the LMF machine readable dictionary (MRD) extension. However these resources would be augmented externally by the semantification mechanism described in the next section.

As we will see later in this paper, derived correspondences between/among the existing lexical resource elements should be organized as a kind of secondary language resource in order to be reused.

3 Semantifying Bilingual Lexical Resource Entries

The semantification of a bilingual lexical resource entry is a necessary substep when associating possibly corresponding lexical concepts in different languages. In principle, the source language (SL) expression (entry word) is first associated with a sense in an SL lexical semantic resource. Then, we seek a possible corresponding sense for the target language (TL) expression (translation equivalent) in a TL lexical semantic resource. This process enriches the bilingual lexical resource by grounding it in the lexical semantic resources in the SL and TL.

3.1 Necessity of Semantification

Bilingual dictionaries provide lexical items in one language with counterparts in another language that are similar in meaning and usage. However, although this definition is fairly straightforward, bilingual dictionaries do exhibit problems that need to be addressed, mainly owing to differences in concept formation in different languages (Svensén 2009). Although the idea of using bilingual lexical resources to integrate semantic resources is not new, as demonstrated by Daudé (1999) or Chen (2002), bilingual dictionaries, in general, have attracted less attention than monolingual dictionaries. As pointed out by Fontenelle (1997), this may, in part, be owing to their less structured machine-readable data format, making it harder for a researcher to mine useful information from bilingual resources. However, a standardized modeling framework such as the ISO LMF can enable more bilingual lexical resources to be disseminated in a well-structured format. The LMF introduces the MRD extension to provide a metamodel to represent monolingual/bilingual dictionaries that are primarily compiled for human use.

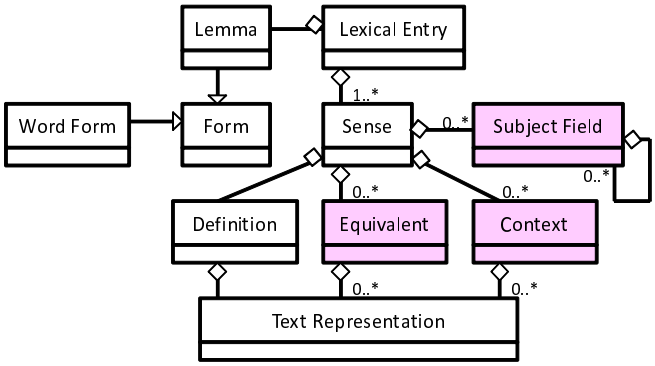

Figure 4: LMF MRD Class Model.

Figure 4 provides an overview of the LMF MRD extension in a UML diagram. It shows that the translation equivalents in the TL for an entry word in the SL are represented by using `Equivalent` nodes, each of which is associated with a `Sense` node of the `Lexical Entry` node. The figure also shows that a translation equivalent is represented by an instance of `Text Representation` class, which basically carries a text string that may be annotated with linguistic data categories. This simple and somewhat unstructured configuration is reasonable and can be acceptable, given the fact that most bilingual resources are structurally messy. However, the configuration may be insufficient if

we are to exploit a bilingual dictionary as a kind of semantic resource and leverage it as a bridge to associate potentially corresponding lexical concepts in different languages. This motivated us to develop a framework to semantify bilingual lexical resources.

3.2 Framework of Semantification

Figure 5 shows the process of semantification. It is noteworthy that before the semantification, the bilingual lexical entry is represented according to the definition in the LMF MRD extension.

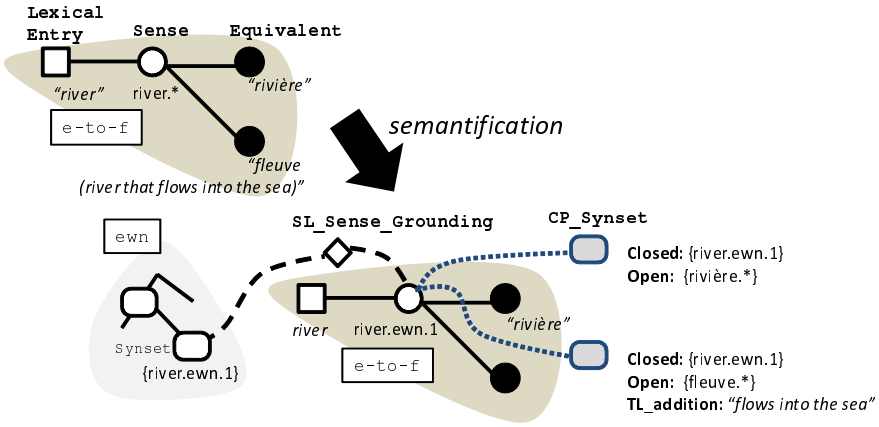

Figure 5: Example of Semantification of a Bilingual Dictionary Entry.

The semantification is as follows:

1. We first perform *SL sense grounding* to associate the `Sense` node in the bilingual lexical resource `e-to-f` with a `Sense` node in the SL lexical semantic resource `ewn`. To accomplish this, a computational lexical semantic matching process first looks for possibly corresponding `Sense` nodes in `ewn`. This process[5], is never decisive, even if it makes full use of the information, such as the entry word itself, a gloss description, or additional semantic markers, provided in the lexical resources. Therefore, a human judgment is then necessary to choose among the candidates and establish a correspondence. Once the correspondence has been established, the formerly underspecified word sense `river.*` in `e-to-f` is disambiguated as `river.ewn.1`. Here `ewn.1` is an identifier[6] of the `Sense` node in `ewn`. At the same time, these two `Sense` nodes are interlinked by an `SL_Sense_Grounding` node, as shown in the Fig. 5.

2. Two `CP_Synset` nodes are then created. For example, the cross-lingual pseudo synset {`river.ewn.1, rivière.*`} is associated with the upper `CP_Synset` node, indicating that the intersection of these two senses denotes a multilingual lexical concept across individual languages. However, note that the sense `rivière.*` indicates that it is not yet grounded to a French lexical semantic resource, and so the `CP_synset` node is still underspecified. In the figure, the set marked **Closed** represents the set of grounded senses, whereas the set marked **Open** denotes the still underspecified senses. These two sets together define the current status of the multilingual pseudo synset. It should be noted that the `Sense` node in the `e-to-f` dictionary is associated with two `CP_Synset` nodes. This is different from the original LMF specification, in which a `Sense` node can only be associated with one `Synset` node. It does not matter, however, as the associations are accomplished only externally, thereby keeping the existing LMF-modeled resource intact.

3. The additional description of the second translation equivalent ”fleuve,” which is a ”river that flows into the sea,” is encoded as the value of the `TL_addition` feature and is stored in the

[5]We are now developing the process, which basically relies on textual overlap (Banerjee and Pedersen 2003).
[6]A rigorous specification has not yet been determined.

`CP_synset` node. As discussed in the next subsection, additional descriptions in a bilingual lexical resource offer useful information to fill the semantic gap between an entry word and the translation equivalents. This information includes semantic restrictions on the translation equivalents, as well as collocational or phrasal equivalents that detail the semantic range of an entry word. However, to extract the information from an additional description, we need to analyze the presented translation equivalent appropriately. This process would be highly resource-dependent, owning to lack of a standardized presentation format. Nevertheless, a technique to extract differentia (O'hara and Wiebe 2004) can be applied, as some of the translation equivalents are given in the so-called *genus-differentia* expression pattern.

4. Although it is not depicted in Fig. 5, if necessary, two underspecified TL senses, will eventually be grounded to the corresponding `Sense` nodes in a French lexical semantic resource. This sub-process is called *TL sense grounding* and is organized in a similar way to that of SL sense grounding, requiring a computational lexical semantic matching process with human intervention. However it may be a more difficult process, because, in general, translation equivalents provided in a bilingual resource are not well structured and tend to lack rich semantic descriptions.

3.3 Dealing with Partial Equivalences

The method used for creating a `CP_Synset` node should consider the nature of the translation equivalents given in a variety of bilingual resources. Translation equivalence can be classified into full equivalence, partial equivalence or zero equivalence (Svensén 2009). He points out that this classification is rough, but important, in the sense that it may determine the way in which a translation equivalent is presented. Among these, partial equivalence is the most noteworthy, because *equivalent differentiation* has to be implemented in the dictionary description in some way, and the relevant information should be extracted and encoded in the computational representation. The cases of partial equivalence can be further divided into *convergence* (*neutralization*) or *divergence*.

The English-to-French correspondences in the motivating example can be classified as an instance of divergence. Another example of divergence is presented by the Japanese word *shujin*, which, in English, corresponds to *host* or *hostess*, depending on the gender of the person[7]. This example can be represented in a similar way to Fig. 5: a `CP_synset` node for {`shujin.jwn.1, host.*`}, with `TL_addition` "male", and another `CP_synset` for {`shujin.jwn.1, hostess.*`}, with `TL_addition` "female." These examples show that in cases of divergence, an SL sense is divided into a set of finer-grained concepts. Generally, a divergence instance is signalled by the additional description that specifies the sense or semantic range of a translation equivalent.

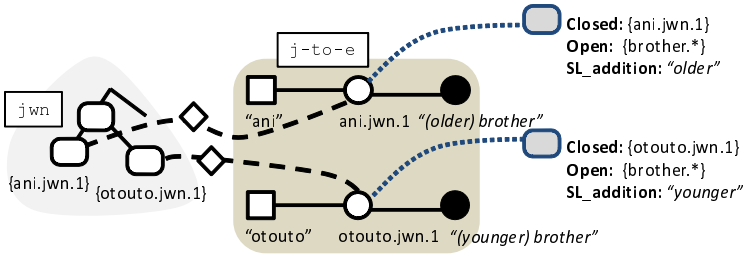

Figure 6: Sample Representation of Conversion-type Partial Equivalence.

Convergence can be illustrated by the example schematized in Figure 6, in which the Japanese word *ani* (elder brother) and *otouto* (younger brother) are jointly associated with the English word *brother*, in the sense of `blood brother`. Contrary to the divergence cases, a convergence instance may be indicated by a phrasal translation equivalent that preserves, or tries to convey, the finer-grained SL meaning.

[7]Actually, the EDR bilingual dictionary (`http://www2.nict.go.jp/r/r312/EDR/`) presents: "$\langle\langle male \rangle\rangle$ host" and "$\langle\langle female \rangle\rangle$ hostess," respectively.

To encode the semantic restriction to the entry word in the SL, we introduce the `SL_addition` feature, as shown in Fig. 6. It should be noted that the two underspecified `CP_synset` nodes would eventually be grounded to the same `Sense` node in an English semantic resource and hence disambiguated and *converged*.

4 Modeling Cross-lingual/Interlingual Correspondences for Reuse

4.1 *Overall Picture*

Figure 7 shows almost the entire representation of the motivating example, providing more detail than the brief sketch shown in Fig. 3. Note that the numbered `CP_Synset` nodes are placed at logically identical positions to those in Fig. 3. In Fig. 7, we introduce instances of the class `TL_Sense_Grounding` (shaded diamonds): a `TL_Sense_Grounding` node is created when the open translation equivalent of an `MP_Synset` node is closed by being grounded to a `Synset` node in the TL lexical semantic resource. With this grounding, together with the `SL_Sense_grounding`, an entry in a bilingual lexical resource works as a bridge from an SL lexical concept to the corresponding TL lexical concept via the `MP_Synset` node.

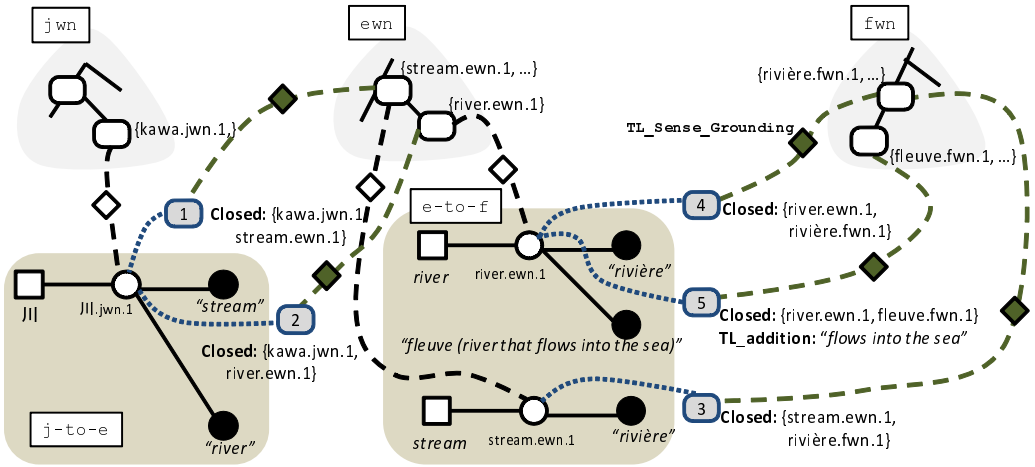

Figure 7: Proposed Representation of the Motivating Example.

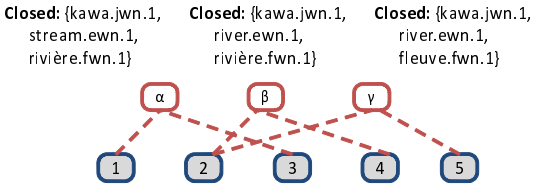

Figure 8: Resulted Lattice-like Structure.

To avoid an unnecessarily complicated diagram, Fig. 8 shows an extra part of the configuration shown in Fig. 7. In this figure, three `MP_Synset` nodes (indicated by Greek letters) are introduced, and linked to the associated `CP_Synset` nodes. At the time of writing this paper, the underlying computational process for deriving the indirect correspondences was still under investigation. However, it is however obvious that the process has to properly filter out inappropriate transitivities to avoid the semantic drift across languages. Again, this would need human intervention, but this may require that the person has competence for all the relevant languages. Therefore an effective machinery to assist him/her to make judgments will be necessary.

Incremental creation of the `MP_Synset` nodes gradually forms a lattice-like multilingual concept structure. This suggests that our proposed framework is similar to SIM*u*LLDA (Janssen 2004), which

applies formal concept analysis (FCA) to derive a concept lattice with the words and formal concepts. However, our framework is clearly different in the sense that we propose an LMF-based representation framework, while considering an incremental formulation of a distributed network structure, as discussed.

4.2 Specifications of the Proposed Constructs

All in all, we have proposed four classes in this paper: `CP_Synset`, `MP_Synset`, `SL_Sense_Grounding`, and `TL_Sense_Grounding`. These classes, which could extend the current ISO LMF, are specified as follows.

- A `CP_Synset` node is initiated when a lexical entry in a bilingual lexical resource is activated.
- An `MP_Synset` node is introduced when `CP_Synset`/`MP_Synset` nodes are combined to define a multilingual pseudo synset.
- An instance node of the `SL_Sense_Grounding` class associates a `Sense` node of an existing bilingual lexical resource entry with the corresponding `Synset` node in an SL lexical semantic resource. In the original LMF, `Sense`-to-`Synset` association is direct and does not require an intermediate node. However, the insertion of an `SL_Sense_Grounding` node is necessary to record the detail of the lexical semantic matching process.
- An instance node of the `TL_Sense_Grounding` class associates the translation equivalent of a bilingual lexical resource entry with the corresponding TL `Synset` node, closing the formerly open translation equivalent.

Central to our framework is the `CP_Synset` and `MP_Synset` classes, which are similar to the LMF `Synset` class in the sense that an instance of these classes represents a set of synonymous senses. However, the `CP_Synset` and `MP_Synset` classes differ from the LMF `Synset` class, because an instance node of the classes gathers synonymous senses across the languages. The LMF `Sense Axis` class is another LMF construct that has something in common with the `MP_Synset` class is. However, we strongly expect that with the `MP_Synset` class, multilingual correspondences will be incrementally recovered and established, while also pointing to the `Sense` nodes in bilingual lexical resources.

4.3 Toward Reusing Recovered Correspondences

Recovered and established cross-lingual/interlingual correspondences should be made persistent somewhere on the Web-based linguistic service infrastructure, so that they can be reused. In other words, these correspondences should be converted into a sort of secondary language resource. Just like the `Sense Axis` class in the original LMF, instances of the `CP_Synset` and `MP_Synset` classes can be aggregated in an instance of the `Lexical Resource`. In this way, the `Lexical Resource` instance can indirectly associate the involved `Lexicon` instances, which are existing primary resources.

However, to make this scenario work, the following issues have to be addressed.

- All the nodes and links external to the existing language resources have to be properly stored somewhere in the infrastructure and made retrievable. This means that standardized Web APIs that enable the search and retrieval of the storage have to be provided.
- At the same time, relevant elements of the existing language resources, such as `Synset` nodes or `Sense` nodes, have to be indexed and be retrievable externally. Assigning global identifiers (URIs) to the elements may be a feasible way to do this. This may also facilitate the servicization of language resources as exemplified in (Savas et al. 2010).

5 Related Work

This paper discusses a framework for representing a global and distributed lexical semantic network, while presupposing an environment in which a number of lexical resources have been Web-servicized. Given such an environment, (Calzolari 2008) has pointed out the possibility of creating new resources on the basis of existing resources, and some work in this direction has been published, such as Soria et al. (2009) and Savas et al. (2010). This line of work is expected to improve further and increase, as Web-based linguistic service infrastructures evolve and gain popularity.

Obviously, another related area of research is lexicon modeling. Although the ISO LMF will undoubtedly be used as a solid and shared framework, requirements to its revisions/extensions continue to emerge. Among them, Maks et al. (2008) pointed out that LMF should more explicitly represent language-dependent usage and contrasts, and they proposes a model that compromises between the MRD extension and the multilingual extension. This solution might be reasonable, if we are to represent an existing bilingual dictionary precisely. Nevertheless, the solution may not be sufficient to model and represent an evolving distributed lexical semantic network, which is a prerequisite for this paper. The problem raised up by Maks et al. (2008) is closely related to the issue posed by Trippel (2010), in which he states: *LMF provides the container for combining such resources of different types, but does not merge them into one formalism*. Given this motivation, he presented a formal lexicon model called *Lexicon Graph*, arguing that the lossless combination of lexical resources could be accomplished.

6 Conclusions

Presupposing a highly servicized language resources environment, this paper proposed a representation framework for cross-lingual/interlingual lexical semantic correspondences that would be recovered incrementally on a Web-based linguistic service infrastructure. The main contribution of this paper is twofold: (1) the notion of *pseudo synset*, which is introduced to represent pseudo lexical concepts shared by more than one language; (2) the framework for *semantifying bilingual lexical resources*, which allows bilingual lexical resources to be used as a bridge to associate lexical concepts in different languages. This paper also discussed how the recovered correspondences can be organized as a dynamic *secondary language resource*, while examining a set of possible extensions to the ISO LMF.

For future work, several items need to be pursued. First we have to extend the representation framework to appropriately accommodate verb and adjective concepts, in which more complicated relationships among linguistic elements have to be organized. Second, we plan to work further on the semantification of bilingual lexical resources. In particular, we intend to devise a formalism and mechanism to represent multi-word lexical entries and complicated translation equivalents. Multi-word expressions are more frequently observed in bilingual resources compared to monolingual resources; they are useful to describe the lexical semantic gaps between the languages. Last but not least, we intend to implement prototype services around some existing lexical resources. To do this, along with the basic semantic matching processes, we have to establish an effective workflow that involves human assessors to approve the recovered cross-lingual correspondences and the inferred multilingual correspondences. In this regard, the notion of a *sense pool* and the verification process proposed by Yu et al. (2007) should be highly relevant as a reference.

Acknowledgments

The author greatly appreciates anonymous reviewers for their thoughtful and informative comments. The presented work supported by KAKENHI (21520401) and the Strategic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communications of Japan.

References

[Banerjee and Pedersen 2003] Banerjee, S., and Pedersen, J. (2003). Extended Gloss Overlaps as a Measure of Semantic Relatedness. In: *IJCAI 2003*, pp.805–810.

[Calzolari 2008] Calzolari, N. (2008). Approaches towards a 'Lexical Web': the Role of Interoperability. In: *ICGL2008*, pp.34–42.

[Chen 2002] Chen, H.H., Lin, C.C., and Lin, W.C. (2002). Building a Chinese-English WordNet for Translingual Application. In: *ACM Transactions on Asian Language Information Processing*, Vol.1, No.2, pp.103–122.

[Daudé 1999] Daudé, J., Padró, L., and Rigau, G. (1999). Mapping Multilingual Hierarchies Using Relaxation Labeling. In: *EMNLP/VLC-99*.

[Fontenelle 1997] Fontenelle, T. (1997). Using a Bilingual Dictionary to Create Semantic Networks. In: *International Journal of Lexicography*, Vol.10, No.4, pp.275–303.

[Francopoulo et al. 2009] Francopoulo, G., Bel, N. et al. (2009). Multilingual Resources for NLP in the Lexical Markup Framework (LMF). In: *Language Resources and Evaluation*, Vol.43, No.1, pp. 57–70.

[Hartmann 2005] Pure or Hybrid? The Development of Mixed Dictionary Genres. In: *Linguistics and Literature*, Vol.3, No.2, pp.192–208.

[Janssen 2004] Janssen, M. (2004). Multilingual Lexical Databases, Lexical Gaps, and SIM*u*LLDA. In: *International Journal of Lexicography*, Vol.17, Vol.2, pp.138–154.

[Maks et al. 2008] Maks, I., Tibeius, C., and van Veenendaal R. (2008). Standardising Bilingual Lexical Resources according to the Lexicon Markup Framework. In: *LREC 2008*, pp.1723–1727.

[O'hara and Wiebe 2004] O'hara, T., and Wiebe, J. (2004). Empirical Acquisition of Differentiating Relations from Definitions. In: *COLING 2004 Workshop on Enhancing and Using Electronic Dictionaries*, pp.77–80.

[Savas et al. 2010] Savas, B., Hayashi, Y., Monachini, M., Soria, C., and Calzolari, N. (2010). An LMF-based Web Service for Accessing WordNet-type Semantic Lexicons. In: *LREC2010*, pp.507–513.

[Soria et al. 2009] Soria, C., Monachini, M., Bertagna, F., Calzolari, N., Huan, C.R., Hsieh, S.K., Marchetti, A., and Tesconi, M. (2009). Exploring Interoperability of Language Resources: The Case of Cross-Lingual Semi-automatic Enrichment of Wordnets. In: *Language Resources and Evaluation*, Vol.43, pp.87–96.

[Svensén 2009] Svensén, B. (2009). Equivalents in Bilingual Dictionaries. In: Svensén, B. *A Handbook of Lexicography*, Cambridge University Press, pp.253–280.

[Trippel 2010] Trippel, T. (2010). Representation Formats and Models for Lexicons. In: Witt, A., and Metzing, D. (Eds.), *Linguistic Modeling of Information and Markup Languages*, Springer, pp.165–184.

[Vossen 2004] Vossen, P. (2004). EuroWordNet: A Multilingual Database of Autonomous and Language-Specific Wordnets Connected via an Inter-lingual Index. In: *International Journal of Lexicography*, Vol.17, No.2, pp.161–173.

[Yu et al. 2007] Yu, L.C., Wu, C.H., Philpet, A., and Hovy, E. (2007). OntoNotes: Sense Pool Verification Using Google N-gram and Statistical Tests. In: *OntoLex 2007*.