How Stable are WordNet Synsets?

Eric Kafe

MegaDoc, Charlottenlund, Denmark,
kafe@megadoc.net

Abstract. The diachronic study of the permanent WordNet *sense keys* reveals that the WordNet synonym sets have stayed very stable through every version of the lexical database since 1.5 (1995), even though the synset identifiers continually changed. In particular, contrary to expectations, 94.5% of the WordNet 1.5 synsets still persisted in the latest 2012 version, compared to only 89.2% of the corresponding sense keys. Meanwhile, the splits and merges between synonym sets remained few and simple. We discuss implications of these results for WordNet mappings, and present tables that allow to estimate the lexicographic effort needed for updating WordNet-based resources to newer WordNet versions.

Keywords: WordNet, Sense Keys, Synsets, Mappings

1 Introduction

1.1 Sense Keys and Synset Offsets

Wordnets cover an increasing number of languages, and interoperate by using identifiers from the Princeton WordNet (PWN) [3] lexical database. PWN groups words that share the same meaning in synonym sets (*synsets*). While the identifier for each synonym set (the *synset offset* [14]) changes between each version of the database, each individual word sense has a stable identifier (the *sense key*), which does not change across different PWN versions. So, according to the WordNet manual, "A *sense key* is the best way to represent a sense in semantic tagging or other systems that refer to WordNet senses" [13].

Since WordNet 1.5SC (1995), sense keys are *unique*: each word sense is a member of one and only one synonym set, so each sense key maps to only one synset offset in a given WordNet version. Additionally, each synonym set contains one and only one sense of each word that share this sense, i. e. each synset offset corresponds to only one sense key of each word.

1.2 Mappings and Updates

However, foreign language wordnets have mostly been mapped to PWN through ever-changing *synset offsets*, and thus bound to one particular version of PWN, which hinders interoperability between wordnets bound to different versions.

Daudé et al. [1] produced a complete set of mappings between all PWN versions that achieve almost perfect recall by a relaxation of precision, but did not use the sense keys as a mapping criterion. Also, updating foreign language wordnets to a newer version of PWN requires additional lexicographic efforts, because the changes (splits, merges, deletions) in the PWN synsets do not always correspond to the composition of the foreign language synonym sets.

So, in order to improve the precision of the mappings when updating between PWN versions, foreign language lexicographers need an accurate picture of the changes that occurred between these versions. But previous analyses have been limited to one PWN source and target pair: WN 1.5-1.6 [1], WN 1.6-3.0 [4], WN 3.0-3.1 [11].

1.3 The Stability of WordNet Identifiers

The present study aims to investigate the stability of the two essential entities of the PWN databases (the word senses and the synonym sets), by tracking their respective identifiers (the sense keys and the synset offsets) across all modern versions, ranging from WordNet 1.5 to the latest WordNet 3.1.1 for SQL (version name suggested by Randee Tengi from the PWN team).

Since the sense keys are unique and persistent, they permit to observe their groupings in synonym sets across PWN versions, and to trace how these synsets evolve in the database over time. Even though synset offsets change between versions, we can follow the sense keys of their members, and obtain an exact recension of all the splits, merges, additions and deletions that occurred between PWN versions, and thus estimate the lexicographic effort needed in order to achieve linguistically satisfying mappings.

2 Methods

2.1 The Sense Key Index

The unique input to our analysis is the *ski-pwn-flat.tab* file from the Sense Key Index (SKI) [7], built from the *index.sense* files included in every PWN version since 1.5. In this form, the SKI is a complete table of tab-separated quadruples (sense key, WordNet version, part of speech, synset offset), linking every sense key to its synset offset in all PWN versions between 1.5 and 3.1.1.

The SKI supports a simple *mapping inference rule*, stating that whenever the same sense key is present in both PWN versions *v1* and *v2*, then a bidirectional mapping link exists between the respective synsets of this key, *s1* and *s2*:

Rule 1: *Sense Key Identity*

$$\frac{\begin{array}{ll} WN_{v1}: & Key_k \in Synset_{s1} \\ WN_{v2}: & Key_k \in Synset_{s2} \end{array}}{Map \quad WN_{v1:s1} \leftrightarrow WN_{v2:s2}} \tag{1}$$

This inference is always valid for *identical sense keys*, so mappings that only use this rule do not produce *false positives*, and have thus 100% precision.

2.2 Analysis

The sense keys After collapsing the *part of speech* and *synset offset* fields from the SKI database file into the 9-digit *synset id* format used in WNprolog [12], we applied the built-in *xtabs* cross-tabulatation function in the *R* statistical environment [9], to obtain a table containing all the PWN versions as columns, all the sense keys as rows, with the *synset id* corresponding to each sense key and each PWN version in the cells, and *0* when the sense key was absent from the corresponding PWN version.

For each pair of consecutive PWN versions (see Table 1), we count the number of sense keys present in either the source version ($\mathbf{WN}_{source}$) or the target version ($\mathbf{WN}_{target}$), or both. Most sense keys *persist* in both versions, and their percentage expresses the *recall* of mappings that use only *Rule 1*. Sense keys that only appear in the source have been *removed* in the target, and those that only appear in the target have been *added* to the source.

The *persistent* and *removed* sense keys add up to $Total_{source}$, so we calculate their ratios as percentages of $Total_{source}$, which add up to 100. The *persistent* and *added* sense keys add up to $Total_{target}$, but their percentages do not add up to 100, because they are ratios of different totals. Both totals are identical to the *Word-Sense Pairs* reported by the WordNet-team [15].

Persistent, Added and Removed Synonym Sets We analyse the evolution of the synonym sets, by considering whether their corresponding sense keys are present in either or both of the source and target PWN versions (see Table 2).

The source synset offsets of persistent sense keys have at least one translation in the target, and are counted as persistent synsets. Source synset offsets that do not have a sense key present in the target correspond to removed synsets, while target synsets that do not have a sense key that was present in the source, have been added in the PWN update.

These figures and their percentages are calculated as for Table 1: the persistent and removed synsets add up to $Total_{source}$, and their percentages add up to 100. The synset totals are identical to those from each corresponding WN Stats[15] manual page. But, because of splits and merges, the number of persistent synsets in the source (i. e. the figure we use here) is not identical to the number in the target, which together with the number of *added* synsets, would add up to $Total_{target}$.

Split and Merged Synsets The synonym sets counted as persistent here satisfy a minimal condition of stability, because they have at least one sense key present in both PWN versions. Extending the previous Rule (1) to synonyms allows to increase *recall*, by mapping removed sense keys to the target synset of their synonyms:

Rule 2: *Persistent Synonymy*

$$\frac{\begin{array}{ll} Map & WN_{v1:s1} \leftrightarrow WN_{v2:s2} \\ WN_{v1}: & Key_k \in Synset_{s1} \end{array}}{\begin{array}{ll} WN_{v2}: & Key_k \in Synset_{s2} \end{array}} \quad (2)$$

Rule 2 applies a mapping link established by Rule 1 to a sensekey *k* from *s1*, to predict that *k* belongs to *s2* in PWN version *v2*. But Rule 2 produces fallacies when *s1* was split into different target synsets, where Rule 1 only holds for some synonyms of *k*, but not for *k* itself.

Studying the evolution of the sensekeys allows us to detect all splits or merges, and to assess their frequency and complexity, i. e. the maximal number of synonym sets involved in one split or merge operation (see Table 3). This allows to precisely identify and count the maximal number of *false positives* that Rule 2 can produce. By contrast, other heuristics like gloss similarity [1] are more uncertain, and therefore not considered in this study.

3 Results

3.1 The Persistence of Word Senses

Table 1 displays the number of persistent, added, and removed sense keys for the nine WordNet updates from version 1.5 to 3.1.1, and four typical long-distance updates between non-consecutive versions, which are relevant for some foreign language wordnets [2, 8], or studied in previous literature [1, 4, 11].

Table 1. Persistence of the SENSE KEYS between WordNet versions

$\mathbf{WN}_{source}$	$\mathbf{WN}_{target}$	$Total_{source}$	$Total_{target}$	*Added* %		*Removed* %		*Persist* %	
1.5	**1.5SC**	168082	170243	8466	5	6305	3.8	161777	96.2
1.5SC	**1.6**	170243	173941	9526	5.5	5828	3.4	164415	96.6
1.6	**1.7**	173941	192460	19799	10.3	1280	0.7	172661	99.3
1.7	**1.7.1**	192460	195817	3652	1.9	295	0.2	192165	99.8
1.7.1	**2.0**	195817	203145	9075	4.5	1747	0.9	194070	99.1
2.0	**2.1**	203145	207016	6164	3	2293	1.1	200852	98.9
2.1	**3.0**	207016	206941	2316	1.1	2391	1.2	204625	98.8
3.0	**3.1**	206941	207235	676	0.3	382	0.2	206559	99.8
3.1	**3.1.1**	207235	206353	39	0	921	0.4	206314	99.6
1.5	**3.1.1**	168082	206353	56431	27.3	18160	10.8	149922	89.2
1.5	**1.6**	168082	173941	17818	10.2	11959	7.1	156123	92.9
1.6	**3.0**	173941	206941	40068	19.4	7068	4.1	166873	95.9
3.0	**3.1.1**	206941	206353	715	0.3	1303	0.6	205638	99.4

Table 1 shows a high persistence of the sense keys after version 1.6: less than 1% were typically removed between consecutive versions, the percentage of *persistent* keys was generally above 99. But before version 1.6, the persistence was a little lower, with approx. 3% removals between versions. For long-distance updates, the lost sense keys accumulate: in total 18160 sense keys have been removed since PWN 1.5, so the ratio of keys from PWN 1.5 that persist in the latest PWN 3.1.1 drops to 89.2%. Most often, the number of additions have by far exceeded the deletions, the only exception being the latest WN 3.1.1 update, which mostly consisted in removals.

3.2 The Persistence of Synonym Sets

Table 2 shows that the synonym sets were always more persistent than the individual sense keys. The lowest persistence rate was 94.5% for the long-distance update from PWN 1.5 to 3.1.1.

Table 2. Persistence of the SYNONYM SETS between WordNet versions

$\mathbf{WN}_{source}$	$\mathbf{WN}_{target}$	$Total_{source}$	$Total_{target}$	*Added*	%	*Removed*	%	*Persist*	%
1.5	**1.5SC**	91581	95137	4597	4.8	1325	1.4	90256	98.6
1.5SC	**1.6**	95137	99642	5649	5.7	1217	1.3	93920	98.7
1.6	**1.7**	99642	109377	9958	9.1	375	0.4	99267	99.6
1.7	**1.7.1**	109377	111223	1921	1.7	112	0.1	109265	99.9
1.7.1	**2.0**	111223	115424	4849	4.2	720	0.6	110503	99.4
2.0	**2.1**	115424	117597	3148	2.7	1012	0.9	114412	99.1
2.1	**3.0**	117597	117659	1155	1	1111	0.9	116486	99.1
3.0	**3.1**	117659	117791	256	0.2	126	0.1	117533	99.9
3.1	**3.1.1**	117791	117371	15	0	436	0.4	117355	99.6
1.5	**3.1.1**	91581	117371	30216	25.7	5048	5.5	86533	94.5
1.5	**1.6**	91581	99642	10196	10.2	2492	2.7	89089	97.3
1.6	**3.0**	99642	117659	20660	17.6	2958	3	96684	97
3.0	**3.1.1**	117659	117371	272	0.2	562	0.5	117097	99.5

This result should actually be expected, considering that removed word senses still can be mapped to target synonym sets through their synonyms. For example, although the adjective sense key for "froward" disappeared between WN 3.1 and 3.1.1 because the orthography of the lemma was corrected to "forward", it is still mapped through synonyms like "headstrong". So mappings that link synset offsets have a higher recall than those that link sense keys, because they cover whole sets of words, and thus avoid some of the losses incurred from the removal of individual sense keys. However, when synsets are split, mapping each key to all its synonyms causes a loss of precision, which we can quantify through a more precise analysis of the splits.

3.3 The Stability of the Synonym Sets

In a mapping with unique pairs of (source , target) synset offsets, split synsets are those appearing more than once in the source column, while merged synsets are those appearing more than once in the target. The number of times that these synsets appear is a measure of the complexity of the split or merge operation. We indicate this size with a subscript, so that $split_2$ and $split_3$ are the number of synsets that were split in respectively two or three different target synsets. Similarly, $merged_2$ and $merged_3$ are the number of merges from two or three different source synsets. Some synonym sets are both split and merged, and we indicate their frequency as $\&^{merged}_{split}$.

After PWN version 1.5SC, $split_2$ and $split_3$ add up to the total number of splits. Similarly, $merged_2$ and $merged_3$ add up to the total number of merges. Thus, between two consecutive WordNet versions after 1.5SC, no source synset was split into more than three target synsets, and no target synset was merged from more than three source synsets. Only in the mapping between WordNet 1.5 and 1.5SC, the total number of splits includes a very small number of four and five-way splits.

Table 3. Splits and Merges in the SYNONYM SETS between WordNet versions

$\mathbf{WN}_{source}$	$\mathbf{WN}_{target}$	*Split*	$split_2$	$split_3$	*Merged*	$merged_2$	$merged_3$	$\&^{merged}_{split}$
1.5	**1.5SC**	489	459	26	232	223	9	142
1.5SC	**1.6**	268	254	14	207	205	2	96
1.6	**1.7**	223	218	5	76	76	0	45
1.7	**1.7.1**	58	57	1	22	22	0	6
1.7.1	**2.0**	128	124	4	60	60	0	30
2.0	**2.1**	93	89	4	60	60	0	22
2.1	**3.0**	85	84	1	66	64	2	27
3.0	**3.1**	33	33	0	31	31	0	11
3.1	**3.1.1**	1	1	0	0	0	0	0
1.5	**3.1.1**	1202	1125	72	649	634	15	359
1.5	**1.6**	733	683	45	421	409	12	236
1.6	**3.0**	559	540	19	260	257	3	124
3.0	**3.1.1**	33	33	0	31	31	0	11

The number and size of the splits and merges was generally low, and there were always more splits than merges. Almost all splits and merges only involved two synsets, and operations involving three synsets were very rare. Between the non-consecutive versions, no merge involved more than three synsets. After WordNet 1.5C, the splits were also limited to two or three synsets

Synsets that were split and merged at the same time most often resulted from the migration of a single sense key to another synset. The following example from PWN 2.1 displays an addition (*medusoid*), a deletion (*medusa#2*), a split (*jellyfish*), and a merge (*medusan*). The deletion of *medusa#2* is implied by the fact that there is already a sense of *medusa* in the target synset.

Sense Key	$\mathbf{WN}_{2.1}$	$\mathbf{WN}_{3.0}$
medusoid%1:05:00::	0	101910252
medusa%1:05:01::	101890584	101910252
medusan%1:05:00::	101891041	101910252
medusa%1:05:02::	101891041	0
jellyfish%1:05:00::	101891041	101910747

The following example shows that the adverb *observably* migrated to its antonym set, during the update from WordNet 2.0 to 2.1. In this case, applying the mapping Rule 2 to its source synonyms *imperceptibly* and *unnoticeably* would aggravate the confusion between synonyms and antonyms, instead of resolving it. To avoid such errors, it is crucial to review all the splits manually.

$\textbf{Sense Key}^{merged}_{split}$	$\mathbf{WN}_{2.0}$	$\mathbf{WN}_{2.1}$
imperceptibly%4:02:00::	400369180	400367415
unnoticeably%4:02:00::	400369180	400367415
observably%4:02:00::	400369180	400367669
noticeably%4:02:00::	400369465	400367669
perceptibly%4:02:00::	400369465	400367669

This example also shows that merges do not produce *false positives*, since the other merged source synset (*perceptibly* and *noticeably*) is only mapped to the correct target.

4 Discussion

4.1 WordNet Synsets Are Very Stable

By simply following the sense keys between WordNet versions, we saw that the synonym sets remained very stable throughout. There was never more than a few hundred split or merged synonym sets between consecutive versions and, after version 1.6, the complexity of these changes was often the lowest possible, because each split or merge almost always involved only two synsets, and never more than three.

Lexicographers can use Tables 1, 2 and 3 to estimate the effort required to update a resource between two PWN versions. For example, when updating to PWN 3.0, a resource that uses PWN 1.6 sense keys and just applies Rule 1 would obtain 100% precision and 95.9% recall (Table 1), which can be improved by a review of the 7068 removed sense keys, as well as the collapsed word senses resulting from the 260 merged synsets (Table 3). The synset-based mappings have higher recall (97% in Table 2), which can be improved by reviewing the same 260 merges, and the part of the 7068 removed sense keys that belong to the

2958 *removed* synsets, while the rest of these 7068 removed sense keys could be *false positives* produced by Rule 2, and need to be reviewed in order to increase *precision*, in addition to the 559 splits from Table 3, which do not affect sense keys.

So these results confirm that "sense keys are the best way to represent a sense" [13], but only by a small margin. Contrary to expectations, synset identifiers provide a reasonable alternative, since the splits between most versions are relatively few and simple. As a consequence, stable synset identifiers like the Inter-Lingual Index (ILI) [10, 11] appear viable.

Practical Application For older projects that were originally mapped to PWN 1.5, like [2, 8], upgrading to PWN 3.1.1 requires to review the intersection of the source data with the 1202 PWN splits reported in Table 3.

On the other hand, updating the wordnets from MCR30-2016 [4] to PWN 3.1 is much easier, since only 33 splits need to be checked. One of these is the following example from PWN, where "Pluto" was moved from the Greek to the Roman "gods of the underworld".

Sense Key	$\mathbf{WN}_{3.0}$	**ILI**	$\mathbf{WN}_{3.1}$
aides%1:18:00::	109570298	i86957	109593427
aidoneus%1:18:00::	109570298	i86957	109593427
hades%1:18:00::	109570298	i86957	109593427
pluto%1:18:00::	109570298	i86957	109593643
dis%1:18:00::	109570522	i86958	109593643
orcus%1:18:00::	109570522	i86958	109593643
dis_pater%1:18:00::	0	i86958	109593643

The Spanish WordNet from MCR30-2016 [4] also includes the involved synsets:

spa-30-09570298-n Aides#n#1, Hades#n#2, Plutón#n#1
spa-30-09570522-n orco#n#2

The ILI 3.1 mapping [5] provides correct identifiers at the synset level, but cannot help in mapping local translations of *Pluto* to their adequate PWN 3.1 synset, so the eventual local splits have to be resolved by local lexicographers. Thus, the Spanish lexicographers need to consider whether *Plutón#n#1* should be moved to the same synset as orco#n#2.

Limitations The present study is limited to only two primary *mapping inference rules*, based on *sense key identity* (1) and *persistent synonymy* (2). Additional mapping links can also be inferred automatically from gloss similarity and other relations, as in [1]. However, since these additional heuristics are more uncertain, they should be studied separately, and applied at a later stage. We find further support for this viewpoint in an analysis of the *lower bounds* for the performance of the many-to-many mappings that result from applying only the two more reliable rules (1) and (2).

4.2 Performance Analysis

The true performance of these mappings lies somewhere above a lower bound that can be calculated by finding the theoretical minimum of the number of correct mapping predictions, and the maximal number of possible fallacies.

Table 4. Worst-Case Mapping Performance

	Mapped	**Not Mapped**
True	$tp = SenseKeys_{Persist}$	$tn = 0$
False	$fp = Mapped - tp$	$fn = SenseKeys \in Synsets_{Removed}$

As reference, we use the imaginary performance of a hypothetic ideal mapping which would be able to map everything accurately, achieving 100% precision and 100% recall. In this ideal situation, there are no true negatives ($tn = 0$), so the sense keys pertaining to the removed synsets from Table 2, which our less ideal mapping cannot map, are false negatives (*fn*).

Only mappings resulting from Rule 1 do not produce false positives (*fp*), while all additional mappings resulting from Rule 2 are potentially false. Thus, only the persistent sense keys from Table 1 are the true positives (*tp*), while all the rest of the mapping could be false positives. In this study, we verified that *fp*+*fn* is equal to the number of $SenseKeys_{removed}$.

Rule 2 produces two kinds of false positives. When synsets are split, a simple one-to-many mapping from a source synset into all its target synsets results in a persistent synonymy relation, where all the words that were synonyms in the source remain synonyms in the target. This may hold for some words, but is not true for all, and can introduce dangerous fallacies, as we saw with the migration of the adverb "observably" to its antonym synset. Hence, all the additional mapping links resulting from split synsets may in theory be false positives (*fp*). Likewise, we also consider as potentially false positives all the removed sense keys that are mapped through their synonyms. However, since these do not necessarily correspond to removals in foreign language wordnets, we may expect the number of *fp* to be stricly lower, in practical use, than the value used here.

So, in this set of values, those that represent correct mappings (*tp* and *tn*) have been set to their theoretical minimum, while the values that concern mapping errors (*fp* and *fn*) are set to their theoretical maximum. Thus, these values allow us to use standard formulas to calculate **lower bounds** for the *precision* and *recall* of the mappings.

These results show, as expected, that applying Rule 2 increases *recall* but deteriorates *precision*. However, after version 1.6, both measures show excellent performance.

This analysis differs from human evaluations by considering the whole PWN dataset, instead of smaller samples, so it provides exact metrics, while human evaluations of limited samples add sample and evaluator biases that can yield

Table 5. Performance *lower bounds* of the mappings between WordNet versions

$\mathbf{WN}_{source}$	$\mathbf{WN}_{target}$	*tp*	*fp*	*fn*	*Precision*	*Recall*
1.5	**1.5SC**	161777	4107	2198	97.5	98.7
1.5SC	**1.6**	164415	4132	1696	97.5	99
1.6	**1.7**	172661	735	545	99.6	99.7
1.7	**1.7.1**	192165	139	156	99.9	99.9
1.7.1	**2.0**	194070	704	1043	99.6	99.5
2.0	**2.1**	200852	693	1600	99.7	99.2
2.1	**3.0**	204625	715	1676	99.7	99.2
3.0	**3.1**	206559	180	202	99.9	99.9
3.1	**3.1.1**	206314	52	869	100	99.6
1.5	**3.1.1**	149922	10057	8103	93.7	94.9
1.5	**1.6**	156123	8097	3862	95.1	97.6
1.6	**3.0**	166873	2499	4569	98.5	97.3
3.0	**3.1.1**	205638	232	1071	99.9	99.5

higher *standard error*, resulting in wider confidence intervals. Larger human evaluations are needed, as well as deeper analyses. Both approaches have complementary merits, and allow meaningful comparisons.

4.3 Comparison with Other Mappings

Daudé 2001 [1] produced a complete mapping from PWN 1.5 to 1.6, by applying a relaxation labelling algorithm, with a set of constraints that involved all semantic relations, and additional heuristics such as gloss similarity. They evaluated the results manually, by applying different constraint sets on samples drawn from the monosemous vs. ambiguous nouns, verbs, adjectives and adverbs (4200 synsets in total), and found 98.8% precision and 98.9% recall for the nouns overall, when using the complete constraint set. In all cases, recall was higher than precision, which is consistent with our results concerning early WordNet versions. However, our Table 5 shows higher precision than recall with the later versions, which suggests that a combined approach could lead to improvements.

HyperDic 2012 [6] used a mixed approach to produce a mapping from PWN 3.0 to 3.1, by combining an all-to-all sense key mapping with additional heuristics, meant to improve recall. The mapping is released under the CC-by 3.0 license, and we found that it strictly included all the results from the simple all-to-all approach and, in particular, that the 33 $split_2$ synsets from Table 3 were split in two. The additional heuristics added 80 synsets, so, if these additional mappings are correct, the mixed approach could produce a modest improvement.

CILI 2016 [11] used sense keys to find that 1796 synsets were modified between WN 3.0 and 3.1. This number, as well as their other figures, differ slightly from our findings, but display similar variations. The authors mapped the changes by hand to the ILI, using a one-to-one strategy, where each synset corresponds to only one ILI identifier. But one-to-one mappings have difficulties with split synsets, and particularly sense key migrations, as we saw previously with the example of *Pluto*, so this approach needs to be complemented by a local review of the split synsets.

5 Conclusion

We followed the sense keys between WordNet versions, and obtained exact figures for the number of *added* and *removed* word senses and synonym sets, as well as the number and complexity of the *split* and *merged* synsets.

We found that the splits and merges between versions were few and simple, and that the synsets have remained very stable throughout. Even though their identifiers are unstable, the synsets were always more persistent than the sense keys, especially in the earlier versions. However the sense keys have the advantage of perfect precision, and have stayed almost as persistent as the synsets after PWN 1.6. So both identifiers provide almost equivalent support for highly accurate mappings between the later WordNet versions: sense keys are still preferrable, but synsets are close.

Then, by relying on the solid baseline provided by the persistent sense keys and synsets, the lexicographic work required to update synset-mapped resources to newer versions of WordNet can essentially be reduced to a manual review of relatively few splits and merges, and a moderate amount of removals.

This study was only possible because PWN offers permanent sense keys, so we may expect that other wordnets with permanent identifiers also enjoy more accurate traceability, leading to enhanced interoperability.

Acknowledgement. This paper benefited from the constructive remarks and suggestions by the anonymous reviewers, and the lively discussion session at the *Challenges for Wordnets* (CfWns) workshop at LDK 2017. Special thanks to the sponsors, organisers and participants of *CfWns 2017*.

References

1. Daudé, J., Padró, L., Rigau, G.: A complete wn1.5 to wn1.6 mapping. In: Proceedings of the NAACL Workshop 'WordNet and Other Lexical Resources: Applications, Extensions and Customizations' (NAACL'2001)., Pittsburg, PA, USA (2001)
2. Dziob, A., Piasecki, M., Maziarz, M., Wieczorek, J., Dobrowolska-Pigo, M.: Towards revised system of verb wordnet relations for polish. In: McCrae, J.P., Bond, F., Buitelaar, P., Cimiano, P., Declerck, T., Gracia, J., Kernerman, I., Ponsoda, E.M., Ordan, N., Piasecki, M. (eds.) Proceedings of the LDK workshops: OntoLex, TIAD and Challenges for Wordnets (2017)

3. Fellbaum, C.: WordNet, An Electronic Lexical Database. MIT Press, Cambridge (1998)
4. Gonzalez-Agirre, A., Laparra, E., Rigau, G.: Multilingual central repository version 3.0: upgrading a very large lexical knowledge base. In: Proceedings of the Sixth International Global WordNet Conference (GWC2012). Matsue, Japan (2012)
5. GWA: ili-map-pwn31.tab. In: Collaborative Inter-Lingual Index (CILI). GitHub, https://www.github.com/globalwordnet/ili, retrieved 2017/04/15 (2017)
6. Kafe, E.: Wordnet mapping. In: HyperDic hyper-dictionary. MegaDoc, http://www.hyperdic.net/en/doc/mapping (2012)
7. Kafe, E.: Sense key index (ski). GitHub, https://www.github.com/ekaf/ski, retrieved 2017/04/25 (2017)
8. Kahusk, N., Vider, K.: The revision history of estonian wordnet. In: McCrae, J.P., Bond, F., Buitelaar, P., Cimiano, P., Declerck, T., Gracia, J., Kernerman, I., Ponsoda, E.M., Ordan, N., Piasecki, M. (eds.) Proceedings of the LDK workshops: OntoLex, TIAD and Challenges for Wordnets (2017)
9. R-team: R version 3.3.3. In: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ (2017)
10. Vossen, P.: EuroWordnet General Document. EWN (2002)
11. Vossen, P., Bond, F., McCrae, J.P.: Toward a truly multilingual global wordnet grid. In: Proceedings of the Eigth International Global WordNet Conference (GWC2016). Bucharest, Romania (2016)
12. WordNet-team: Prologdb(5wn) manual page. In: WordNet manual. Princeton University, http://wordnet.princeton.edu/man/prologdb.5WN.html (2010)
13. WordNet-team: Senseidx(5wn) manual page. In: WordNet manual. Princeton University, http://wordnet.princeton.edu/wordnet/man/senseidx.5WN.html (2010)
14. WordNet-team: Wndb(5wn) manual page. In: WordNet manual. Princeton University, http://wordnet.princeton.edu/wordnet/man/wndb.5WN.html (2010)
15. WordNet-team: Wnstats(7wn) manual page. In: WordNet manual. Princeton University, http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html (2010)