Normalizing the translation of idioms of mismatching polarity

L.M. TOVENA*
Centre for Cognitive Science
University of Edinburgh
2 Buccleuch Place
Edinburgh EH8 9LW — Scotland
lt@cogsci.ed.ac.uk

Abstract

The analysis presented in this paper makes it possible to establish stable bidirectional pairs of corresponding expressions for the translation of idioms cross-linguistically mismatching in polarity sensitivity. The strong similarity of cross-linguistic mismatches in polarity with cases of diverging lexicalization patterns is discussed. The meaning of these idiomatic expressions is decomposed into several distinct facets, and particular attention is given to two components, called "negative polarity" and "basic". The possibility of different patterns of lexicalization for these facets is shown to be a cause of instability in the translating pairs, but the separate identification of the facets is then exploited to produce a solution to the issue of the directionality of the translating pairs. We sketch how such analysis could be used during the transfer phase in a machine translation system.

El análisis presentado en este trabajo permite establecer pares de expresiones correspondientes que son estables y bidireccionales, en la traducción de "frases hechas" en que hay diferencias de sensibilidad polar en distintas lenguas. Empezamos demostrando la similitud entre ejemplos de diferencias de sensibilidad polar en distintas lenguas y patrones de lexicalización divergentes. Después, diseccionamos el significado de estas expresiones en varias facetas, y nos centramos en dos componentes, que llamamos "polaridad negativa" y "básico". Por un lado, demostramos que la existencia de los diferentes patrones de lexicalización para estas facetas puede hacer inestables los pares de expresiones correspondientes. Por otro lado,

*The research reported in this paper was supported by the Fonds National Suisse de la Recherche Scientifiq via a *bourse de relève*, which is gratefully acknowledged. Most of the material was presented at the 1993 Worksh on Negation in Edinburgh. I wish to thank Graham Russell and Mercè Prat Sala for comments, and Mercè togeth with Enric Vallduví for translating the abstract into Spanish.

utilizamos la identificación de las facetas para proveer una solución al problema de la direccionalidad de alguno pares. Este análisis nos permite esbozar un tratamiento de tales pares de expresiones durante la fase de *transfer* en los sistemas de traducción automática.

Topic area: machine translation, semantics, negative polarity, idioms

Topic area: traducción automática, semántica, polaridad negativa, frases hechas

1 Introduction

This paper discusses a treatment of cross-linguistic mismatches in polarity sensitivity which makes it possible to establish stable bidirectional pairs of corresponding expressions. We propose an analysis of the meaning of negative polarity idioms in terms of lexical semantics. Cases of mismatches in polarity sensitivity between corresponding expressions across languages are treated as being analogous to cases of diverging lexicalization patterns similar in spirit to (Talmy 1985). The individual encoding of independent facets of meaning of each expression offers a solution to the question of the instability of potential translation pairs. The solution is semantically motivated and language-pair independent. First, we will identify the components of the meaning that are assumed to have different distributions. Then, the different patterns of distribution will be defined, with examples from translation pairs in Italian and English. Finally, the direct use of our analysis in a machine translation system based on transfer will be outlined.

Translation has been viewed quite often as a process of transposing a *message*, generally in written form, from a language, the source language, into another, the target language. Idioms are a widespread phenomenon in natural language. They constitute a problem for translation in general in as much as they are complex expressions whose behaviour and distribution are not consistent across languages. The case in which an idiom in language X does not correspond to an idiom in language Y constitutes a difficulty for translation. Machine translation carries with itself the additional problem of the recognition of idioms. Hence, the encoding of these expressions and the possible lack of matches are among the issues that translation systems must tackle. There is a subset of idioms, e.g. *lift a finger* or *hold a candle to*, which present the additional complication of being of negative polarity. Broadly speaking, their distribution is restricted by positive or negative features of the sentences where they occur. A mismatch in polarity between these expressions in

language X and Y makes the translation even more complex.

A translation pair of expressions is established on the basis of the information content they carry. Potentially, such a pair is not mono-directional, which is to say that it does not presuppose particular instantiations of the source and target languages. The case where both elements of a pair are of negative polarity might be reduced to the standard case of bidirectional pairs, with the proviso that both source and target grammars must check the satisfaction of licensing conditions. It may happen, however, that the expressions selected do not share the same sensitivity to polarity. This implies that the environment in which they can occur may have to satisfy different requirements. For the sake of simplicity, in this paper we adopt the idea that negative polarity expressions have a distribution more constrained than non-sensitive ones. Then, it is easy to see that mismatches in polarity make translation pairs unstable because, if it is true that a move from a more constrained to a less constrained situation seems always possible, the reverse does not necessarily follow. For instance, when the source expression is not polarity sensitive, and the source sentence is not a suitable context for a polarity expression, it seems likely to expect that the target sentence will also be unsuitable, and therefore a target expression of negative polarity will be unacceptable in this case. However, the occasional rejection of the correspondence does not question the existence of the translation pair, nor its potential bidirectionality.[1] Moreover, as shown in (Tovena 1992), despite the presence of mismatches in polarity, some pairs are stable i.e. always bidirectional, while others are not, and cannot be enforced systematically in both directions. Still, allowing only monodirectional pairs for these expressions is too radical a decision because it rules out many suitable translations.

This paper proposes a way of normalizing the pairs, i.e. a way of making them all bidirectional The treatment proposed is independent from the pair of languages considered, i.e. it does require encoding information whose presence is motivated just by the selection of source target languages. Once the lexical choice has been performed, the monolingual grammar takes care of the syntactic structure of the target sentence and, where needed, enforces the requirement of negative polarity licensing without endangering the message.

[1] For instance, the French *bois* and its English translation *wood* constitute a bidirectional pair. However, because of their slightly different denotations, sometimes *timber* must be selected instead.

2 Facets of meaning

The semantic analysis of polarity phenomena proposed by Ladusaw (Ladusaw 1979) relates sentences containing negative polarity items to sets of propositions. There, the notion of scalar endpoint, cf. (Fauconnier 1975), is crucial. Fauconnier proposed to interpret the function of the negative polarity item *any* as being the indication of a low point on an arbitrary scale. In Ladusaw's proposal, semantic contexts that license inferences are identified with the elements in whose scope the members of the scale are contained. For example, let us consider the sentence in (1).

(1) Chloe read ten letters.

A pragmatic scale ranging from an unspecified top to 1 along the dimension of quantity of letters is assumed. This scale is associated with a propositional schema *y reads x*, where *x* is the variable in which we are interested.

(2) ...

↓

Chloe read 20 letters.

Chloe read 10 letters.

Chloe read 5 letters.

Chloe read 1 letter.

For the sake of simplicity, the variable *y* is replaced by the constant *Chloe*. Then, we can further abstract and generalize by reformulating the propositional schema as P(*x*), where P stands for *Chloe reads ... letter(s)*, and *x* is the number of letters that is to be inserted in place of the dots, i.e. it is the scalar expression. According to such a definition, our example can now be rewritten as P(10). Following the gricean maxims of conversation, if the proposition P is true for the element *10*, then it is true for all the elements on the scale from that position upwards, e.g. for *5* and *1*, which is to say that P(10) → P(5) holds, see (Levinson 1983, ch. 3) on scalar implicatures. As usual, the logical negation of the proposition causes the reversal of the direction of the implications on the scale, which is to say that ¬ P(5) → ¬ P(10) holds. Similarly, in natural language the negation of P(1) implies the negation of all the positions on the scale from there downwards, see (3).

(3) ...

↑

Chloe did not read 10 letters.

Chloe did not read 5 letters.

Chloe did not read 1 letter.

A negative polarity item, like *any*, denotes the position at the bottom of a scale where the direction of the implications is reversed, i.e. it runs downwards. An indefinite expression like *a word*, that Bolinger[2] called a *minimizer*, has also been analysed as a negative polarity item. According to such analysis, it is associated with a scale similar to (4).

(4) ...

↑

Chloe did not say three words.

Chloe did not say two words.

Chloe did not say a word.

The scale in (4) is the standard, default scale that represents the interpretation associated with the predictable relation between the expressions *a word, two words*, etc.[3] Now, when interpreting sentence (5), because of the idiomatic reading of the expression *lift a finger*, we associate it with a scale similar to the one presented in (6). Negative polarity items like *lift a finger* are idioms because, for a complex expression to function like a negative polarity item it must be possible to assign the entire expression a single semantic property. Once this has been done, it is no longer possible to divide up the unit and still retain the property.

(5) Chloe did not lift a finger.

---

[2]We quote from Horn (1989).

[3]The presence of the indefinite *a* instead of the numeral *one* is to be linked with the non-referential properties of negative polarity items.

(6) ...

↑

Chloe did not help enormously.

Chloe did not give a helping hand.

Chloe did not lift a finger.

In this scale the alternatives to *lift a finger* are "acts of labour". Strictly speaking, however, sentence (5) could have also been associated with the scale in (7), that expresses a scalar implicature as well, but which does not seem to represent the interpretation in which we are interested. What seems of interest in the case of (5), is not the predictable relation between the expressions *lift a finger, lift two fingers*, etc., but the unpredictable association of *lift a finger* with other words. The particular interpretation of *lift a finger*, more precisely its idiomatic reading, motivates this association.

(7) ...

↑

Chloe did not lift a hand.

...

Chloe did not lift two fingers.

Chloe did not lift a finger.

The scale in (7) represents the scale produced by default, but here the default interpretation is overrided by a more specific interpretation, which is preferred. Such blocking of the application of a default rule by a more specific one, is quite a common phenomenon in the lexicon. The occurrence of this blocking effect with respect to the identification of the interpretation of idioms suggests that idioms are a lexical phenomenon.

These data support the hypothesis that the meaning of expressions like *lift a finger* is the result of the contribution of several components. Negative polarity is a facet of the meaning of this group of idioms. It can be characterized as the property of denoting intensionally scalar endpoints. Every expression comes with a set of alternatives which is a totally or partially ordered set with a unique bottom element.[4] This property alone does not capture the whole meaning of these expressions.

[4]See (Krifka 1989) and references therein for a discussion on the partial/total order of the set.

At least the idiomatic component must be considered too. We call this component the 'basic' one because it constitutes the basis on which the translation pair is established. It also determines the content of the scale, and it is 'idiomatic' because the elements of the ordered set are given by the lexical knowledge of the language user, cf. (Krifka 1989).

3 Diverging patterns of lexicalization

Evidence of the detachability of the two components of the meaning, the negative polarity and the basic – which here corresponds to the idiomatic one – comes from (Tovena 1992). There, we examined cross-linguistic mismatches in negative polarity. It was shown that negative polarity sensitivity is a semantic property whose contribution cannot be ignored in computing the meaning of an expression. As mentioned in the introduction, if there is a mismatch in sensitivity, one would expect to have problems in enforcing the pair in some cases, always the same cases for all pairs, namely when the non-negative polarity source expression is used in non-licensing contexts. Polarity negative expressions have a more restricted distribution, so, if the source sentence is not a suitable context for such an expression, it seems likely to expect that the target sentence will also be unsuitable. Our previous study showed this not to be the case. It turned out that some pairs of expressions are stable and bidirectional, while others are not stable, and cannot be enforced systematically in both directions. Re-examining these data in the light of the analysis of meaning presented above, it appears that cross-linguistic mismatches in polarity may or may not be problematic depending on the distribution of the message over the lexical components of the sentence. In other words, cross-linguistic mismatches in polarity sensitivity can be considered analogous to diverging lexicalization patterns of the components of the meaning we have identified

3.1 Unstable pairs: a case of underlexicalization

As far as idioms of contrasting polarity sensitivity are concerned, a translation pair is unsta when one of the two expressions lexicalizes only one component of the meaning, the basic and the other is contributed by other elements of the sentence. In cases where that exte contribution is missing, the use of the pair will lead to an unacceptable drift in the transla The expressions *hold a candle to* and *reggere il confronto* are an example of an unstable pai

fact, the Italian expression *reggere il confronto* expresses the idea of standing in a comparison, but it is neutral with respect to the position on the scale of comparisons. On the contrary, the English idiom *hold a candle to* is of negative polarity, and can be used only to express unfavourable comparison. Therefore, the Italian expression can be translated by its English counterpart only in sentences expressing unfavourable comparison, as in (8)[5].

(8)a. Dubito che il mio lavoro regga il confronto con il suo.
doubt-1st that my work holds the comparison with hers
b. I doubt that my work holds a candle to hers.

However, when the sentence denotes a higher position on the scale of comparisons, i.e. a favourable one, the translator is forced to select another target expression, as shown in the source and target pair in (9a) and (9b) respectively. Sentences (9c) and (9d) have been included to show that no adjustment is possible. In (9c) the polarity idiom is not licensed, while in (9d) the insertion of the licenser *not* has turned the original message into its opposite.

(9)a. E' chiaro che il mio lavoro regge il confronto con il suo.
it is clear that my work holds the comparison with hers
b. It is obvious that my work stands comparison with hers.
c. # It is obvious that my work holds a candle to hers.
d. # It is obvious that my work does not hold a candle to hers.

3.2 Stable pairs: a case of overlexicalization

Sentences (10) and (11) present an example of a stable pair. The negative polarity Italian expression *sapere che pesci pigliare* can always be translated in English as *be at a loss* and vice versa.

(10)a. Marco è un incapace, dubito che sappia che pesci pigliare.
Mark is useless doubt-1st that know-3rd which fish to catch
b. Mark is useless, I think he is at a loss.

[5] All examples of this section are from (Tovena 1992). The # sign marks indistinctively well/ill-formed unacceptable translations.

(11)a. I do not doubt that he was at a loss in such a predicament.

b. Sono sicura che non sapesse che pesci pigliare in quella brutta situazione.
am sure that not know-3rd which fish to catch in that ugly situation

We propose that the stability of this type of pair rests on the fact that both components of the meaning — i.e. the one defining the content of the scale and the one fixing the position in it — are present in both expressions. The two expressions denote the bottom of a scale of *available resources*.

3.3 Some considerations on stable pairs

As mentioned above, the Italian expression *sapere che pesci pigliare* and its English correspondent *be at a loss* share the same two facets of meaning we have identified as negative polarity and basic. However, if we consider negative polarity to be a semantic phenomenon, we would expect to find a difference in meaning between the two expressions, if anything because they do not share the same sensitivity. As a matter of fact, some adjustments are needed to make the translations acceptable. The verb *dubito* in (10a) is changed into *think* in the translation in (10b), whereas *doubt* in (11a) is changed into *sono sicura* in the translation in (11b). Example (12) shows the modification very clearly; (12a) is correctly translated in (12c). Sentence (12b) reverses the meaning of the original.

(12)a. Dubito che sapesse che pesci pigliare.
doubt-1st that know-3rd which fish to catch
b. # I doubt he was at a loss.

c. I think he was at a loss.

From example (12), it appears that the two expressions under consideration have to be combined with a different number of scalar reverser operators to get equivalent sentences. More precisely, (12c), which contains the non-sensitive expression, has one reverser less. We seem to reach the same conclusion when considering the scales associated with the idiomatic quantified readings of the two expressions, presented respectively in (13) and (14). As (14) shows, the expression *be at a loss* expresses negated existential quantification over amounts of resources. The scalar reverser operator applied to every other instance of the propositional schema with which the scale is associated is not applied externally to *be at a loss*.

(13) ...

↑

Marco non aveva tantissime risorse.

...

Marco non sapeva che pesci pigliare.

(14) ...

↑

Mark did not have plenty of resources.

...

Mark was at a loss.

There is no need for the presence of such an operator for the expression *be at a loss* to be the strongest expression in such a scale. It appears as if the role of the reverser, or the reverser itself, had been incorporated in the expression; in other words, as if it had been lexicalized within that expression. For this same reason it is impossible to build an upward entailing scale containing such an expression, cf. (15).

(15) ...

↓

Mark had plenty of resources.

...

Mark had some resources.

↓

Mark was at a loss.

The null set is not entailed by other quantifiers. Upward entailing scales do not contain an empty set at the bottom, see (Hirschberg 1985) for a discussion. The expression *be at a loss* is the lexicalization — without overt negation — of the empty set of amounts of resources. Therefore, whenever this expression appears in a scale, the empty set is *ipso facto* included in the scale.

4 Monitoring the association of content with form

In the lexicon of a machine translation system, the meaning of these polarity idioms is represente by the facets we have identified above, and is encoded as semantic features[6]. Their non-polari sensitive correspondents carry one or both features according to their own semantics. On t other hand, the property of polarity sensitivity is to be expressed also by a syntactic feature the whole idiom.[7] This makes it possible to pin down the difference in polarity between *lift a fing* and *be at a loss*, and hence the special requirements on the contexts, and still say that in bo cases the expressions are scalar endpoints. No need to say that the expression *be at a loss* will marked semantically for the presence of the scalar reverser. The syntactic label is attached to head of the negative polarity idiom, characterized by its special idiomatic requirements. This h could be viewed as a function taking particular words as its arguments, potentially disregard their semantic contribution, and returning the meaning of the idiom.[8] [9] Such a syntactic lab interpreted by the monolingual grammar, and forces the checking of licensing requirements.

The encoding of the different facets of meaning as independent features contributing to intermediate representation of the translation system makes it possible to have bidirectional tr lation pairs for expressions of mismatching polarity. The process of lexical selection can be vi as being performed directly by adopting pairs of expressions, and controlled indirectly by mapping between the two intermediate representations. More precisely, the basic meaning c used to trigger the selection of optional translation pairs. The set of potential pairs contain the lexicon is trimmed on the basis of the capacity of each target candidate to contribute t intermediate representation without introducing extra semantic features. For the mapping t place, all the components of the meaning of the source intermediate representation are to be in the target one. When present in the source, the semantic polarity feature will also find i in the target. The possible adjustments of the linguistic context described above take pl

[6]The complex and more general question of the proper encoding of idioms is outside the scope of the paper.

[7]This proposal is not necessarily at odds with the semantic analysis adopted in this paper. Rather, it motivation in the relevance of context.

[8]This functional approach is a modification of the analysis in (Gazdar et al. 1985).

[9]As argued in (Tovena 1993), the option of considering licensers as elements of the idiom leads to serious in controlling both their characterization and the stability of the overall meaning of the expression. Hc was noted that a straight characterization of these idioms as negative polarity items introduces a relation — proper of the licensing approach — that clashes with the traditional idea of self-containment of idiom

a global balancing of the two representations reflected in the lexical selection. No extra content can be added. The selection between surviving lexical options, if any, is a matter of translation policy.[10]

The target language grammar operates with the content of the representation, and any commitments to particular expressions neither need to nor can be retracted. The satisfaction of licensing requirements influences the syntactic form, but can no longer alter the message conveyed by the source text. This is possible because the message is monitored during the process of transfer by using information which is motivated independently from the pair of languages involved, contrary to what was required in (Tovena 1992). In fact, the idiomatic negative polarity item is not marked for the possibility of adding a scale reverser in order to provide the needed licenser. Such a modification is understood because of the semantic balance between expressions of mismatching polarity.

Another positive side-effect of our treatment which is valid for machine translation, but is also relevant to natural language processing in general, and perhaps specifically to generation, is the following: the decomposition of the meaning as it has been proposed, allows the possibility of linking the selection of polarity idioms to the process of selection of content words. One is no longer obliged to wait for the full syntactic realization of the sentence in order to know whether it is possible to use certain expressions that, because of their particular requirements, could lead to an unacceptable alteration of the message conveyed. The lexical selection is based on the different components of the meaning. The *massage* to the structure of the sentence sometimes needed to enforce licensing conditions will no longer affect the message conveyed.

5 Conclusions

In this paper, we have identified several components of the meaning of negative polarity idioms. We have termed them *negative polarity* and *basic* components, and we have defined their characteristics. The possibility of different patterns of lexicalization for these facets of meaning is the cause of partial or complete bidirectionality in pairs of corresponding expressions of mismatching polarity sensitivity.

[10] For instance, it is stylistically motivated to prefer an idiom in the target sentence if an idiom is used in the source.

We have shown that unstable translation pairs correspond to cases where only part of the meaning of the polarity idiom has been lexicalized in the corresponding expression. On the other hand, in the case of stable pairs both facets of meaning of the polarity idiom are lexicalized in the corresponding expression. In this latter case, the difference in polarity sensitivity has been connected with the fact that the corresponding non-polarity sensitive expression incorporates also extra material, namely the scale reverser operator.

By checking that the two intermediate sentence representations share the same components of the meaning, it is possible, on the one hand to avoid unacceptable drifts in translation, and on the other to escape the unjustified constraint of adopting in the target sentence the same patterns of lexicalization present in the source.

6 References

Fauconnier G., 1975, Pragmatic Scales and Logical Structure, in *Linguistic Inquiry*, Vol.6, pp. 353-375

Gazdar G., E. Klein, G. Pullum, I. Sag, 1985, *Generalized Phrase Structure Grammar*, Basil Blackwell, Oxford

Hirschberg J.B., 1985, *A Theory of Scalar Implicature*, Doctoral dissertation, Pennsylvania State University, distributed as Technical Report MS-CIS-85-56, Department of Computer and Information Science

Horn L., 1989, *A Natural History of Negation*, Chicago University Press, Chicago

Krifka M., 1989, Polarity Phenomena and Alternative Semantics, in Stokhof M. and L. Torenvliet (eds.), *Proceedings of the Seventh Amsterdam Colloquium*, ITLI, Amsterdam, pp. 277-301

Ladusaw W., 1979, *Polarity sensitivity as inherent scope relations*, Doctoral dissertation, The University of Texas at Austin, published by Garland Publishing Inc., 1980

Levinson S.C., 1983, *Pragmatics*, Cambridge Textbooks in Linguistics, Cambridge University Press, Cambridge

Talmy L., 1985, Lexicalization patterns: semantic structure in lexical forms, in Shopen T. (ed.) *Language typology and syntactic description*, Cambridge University Press, Cambridge

Tovena L.M., 1992, Polarity sensitive idioms in translation, in *Proceeding of the IDIOMS Conference*, ITK Proceedings, Tilburg, pp. 127–140

Tovena L.M., 1993, Negative Polarity and the Definition of Idioms, ms. University of Edinburgh