

CLUE Guidelines

Anabela Barreiro, Luísa Coheur
Tiago Luís, Ângela Costa, and João Graça

name.surname@l2f.inesc-id.pt

June 19, 2012

The CLUE (Cross-Language Unit Elicitation) Guidelines summarize the most important cross-language alignment recommendations, which were collected after aligning bilingual texts between all combinations of English, Spanish, French, and Portuguese of the common test set of the Europarl corpus. In addition to the conventional word alignments, this document includes a succinct set of linguistically-informed and motivated annotation guidelines for cross-language unit alignment that take into account multiword and translation units (including paraphrases). Following these guidelines will help linguistic annotators be more efficient and consistent in aligning special types of selected linguistic phenomena, which are presented in this document with examples.

Contents

1	Annotation guidelines	4
2	General guidelines	4
2.1	Incorrect translation, incorrect word use and typo	4
2.1.1	Incorrect translation	4
2.1.2	Incorrect word use	5
2.1.3	Typo	6
2.2	Incomplete translation or non-translation	6
2.2.1	Incomplete translation	6
2.2.2	Non-translation	7
2.3	Emphatic linguistic structure	7
2.3.1	Tautology and pleonasm	7
2.3.2	Repetition of words or phrases/expressions	8
2.3.3	Additional and missing information	8
2.4	Approximate numeric correspondence	9
2.5	Mismatching pronoun and determiner	10
2.6	Abbreviation versus full word	10
2.7	Punctuation	10
2.7.1	Discrepancy between punctuation marks	10
2.7.2	Optional versus obligatory punctuation	11
2.7.3	Missing and misplaced punctuation mark	12
2.7.4	Comma + coordinating conjunction <i>and</i>	13
2.7.5	Comma + relative pronoun <i>which</i>	13
3	Multiword units	14
3.1	Support verb constructions	14
3.2	Compounds	16
3.3	Phrasal verbs	19
3.4	Prepositional predicates	19
3.4.1	Prepositional verbs	19
3.4.2	Prepositional nouns	20
3.4.3	Prepositional adjectives	21
3.5	Named entities	22
3.6	Date and time expressions	22
3.7	Lexical bundles	23
3.8	Idiomatic expressions	23
3.9	Domain terms	24
3.10	Other expressions	24
4	Lexical versus non-lexical realization	25
4.1	Determiners and zero determiners	25
4.2	Pronoun-dropping phenomena	26
4.2.1	Subject pronoun drop	26
4.2.2	Empty relative pronoun	27

4.3	Relatives versus participial adjectives	27
5	Other linguistic phenomena	28
5.1	Free noun adjuncts	28
5.2	Contracted forms	29
5.3	Singular versus plural (related to determiner)	29
5.4	Active versus passive	30
5.5	Coordination	31
5.6	Noun pre-modifiers	32
5.7	Anaphoric reference	33
5.8	Antonyms and negation constructions	33
5.9	Flexible/loose paraphrasing constructions	34
5.10	Different parts-of-speech with same semantics	34
5.10.1	Verb versus noun predicates (process nouns)	34
5.10.2	[of ADJ N] versus [ADV ADJ]	35
5.10.3	[ADJ N] versus [ADJ Prep N]	35
5.10.4	Gerundive and process nouns	35
5.10.5	[to-V + N] versus [(Prep +) N Prep N]	36
5.11	Impersonal constructions	37
5.12	Romance languages double negation (+ coordination)	37
6	Idiosyncrasies of languages	37
6.1	Portuguese inflected infinitive (peculiar verb tense)	37
6.2	English infinitive (<i>to</i> + V)	39
6.3	French negation (<i>ne pas</i>)	39
6.4	English apostrophe	40
6.5	Focus constructions	40
6.6	Sociolinguistic differences: register and forms of courtesy	41
6.6.1	Addressing	41
6.6.2	Thanking	42

1 Annotation guidelines

Annotation guidelines correspond to the decisions made for the process of alignment of as many segments as possible between two parallel sentences. These alignments correspond to meaning and translation units, represented graphically by the intersection of single segments or blocks. If a word, expression, or phrase in the sentence of one language corresponds semantically to an identical word, expression, or phrase in the sentence of the other language, that should be aligned in a sure (S) or in a possible (P) alignment, depending on this word, expression or phrase being non-ambiguous or ambiguous. If a word, expression, or phrase in the sentence of one language does not correspond semantically to any word, expression, or phrase in the sentence of the other language, no alignment should be made. General and language-specific morphological and syntactic constraints have also been taken into consideration in the alignment process to account for the universality or the variability of a particular language or group of languages when these are in contrast, with the concern for the linguistic structures or grammar of each of these languages.

This document is organized as follows: in Section 2 we describe the general guidelines, in Section 3 we focus on the alignment of multiword units, in Section 4 we discuss lexical and non-lexical realizations and, in Section 5, we explain how to align other linguistic phenomena. Finally, in Section 6 we target idiosyncrasies of language.

2 General guidelines

This section illustrates some general annotation guidelines that have been defined in previous referenced works and presents new ones established on the basis of linguistic knowledge.

2.1 Incorrect translation, incorrect word use and typo

Words that are incorrectly translated should be left unaligned. Similarly, if a multiword unit is mistranslated, contains a word that is being used incorrectly or has a typo, this multiword unit is not block-aligned and none of its internal words are segment-aligned.

2.1.1 Incorrect translation

The following examples contain occurrences of incorrect translations.

EN – both of which are very active
ES – ambos ***están** muy activos
FR – ils sont tous deux très actifs

PT – ***cuja execução se está a revelar positiva**

Action: no segment-alignment of the Spanish expression, which contains an incorrect word (**están* instead of *son*, or instead of no verb)

Action: no segment-alignment of the Portuguese expression, which is not semantically equivalent to any of the expressions in English, Spanish and French

EN – *the vision group has produced* **an excellent model here**

ES – *el vision group* ***tiene** *ha producido un documento muy bueno*

FR – *le vision group a conçu* **un très bon modèle**

PT – *a vision group fez* **uma proposta muito positiva**

Action: no alignment of the Spanish verb *tiene*; the correct verb (compound) comes on its right-hand side *ha producido*

Action: P-block-alignment of the adjective modifier *excellent*, with the Spanish *muy bueno* and the French *très bon*, but no alignment with the Portuguese *muito positiva*

Action: S-segment-alignment of the semantically equivalent individual words of the object noun phrase *model* in English and *modèle* in French

Action: no segment-alignment of the individual words of the object noun phrase *documento* in Spanish and *proposta* in Portuguese

2.1.2 Incorrect word use

The following example contains the occurrence of incorrect word use.

EN – *we believe that*

ES – *creemos que*

FR – *nous pensons que*

PT – *somos* ***que** *opinião que*

Action: S-block-alignment of the English, Spanish and French lexical bundles

Action: no block-alignment of the Portuguese lexical bundle, which contains an incorrect word (**que*, instead of *de*)

2.1.3 Typo

The following examples illustrate occurrences of word typos.

EN – *european investment* ***fun**
ES – *fondo europeo de inversiones*
FR – *fonds européen d’investissement*
PT – *fundo europeu de investimento*

Action: S-block-alignment of the Spanish, French and Portuguese named entities, and subsequent S-segment-alignment of the individual words of these multiword units

Action: no S-block-alignment of the English named entity due to the typo in the word *fund*, with subsequent no S-segment-alignment of its individual words

EN – [...] *we would not be serving the objective of the regulation*
ES – [...] *estaríamos prestando un flaco servicio al objetivo de la norma*
FR – [...] *nous rendrions un maigre service à l’objectif de la norme*
PT – [...] *não estaríamos a cumprir o ***objectivos** prosseguido*

Action: no P-segment-alignment of the the Portuguese plural noun (*objectivos*), which should be in the singular form (*objetivo*)

2.2 Incomplete translation or non-translation

An incomplete translation or non-translation corresponds to a translation in which a word, expression or part of a sentence is missing in the target language. The missing part of the sentence is not aligned.

2.2.1 Incomplete translation

The example below shows an incomplete translation in Portuguese.

EN – *web of rules and regulations*
ES – *jungla de normas, formularios y ventanillas*
FR – *jungle des règles, formulaires et guichets*
PT – *floresta de regras* \emptyset

Action: S-block-alignment of the English idiomatic phrase *web of rules* with the its idiomatic equivalents *jungla de normas* in Spanish, *jungle des règles* in French, and *floresta de regras* in Portuguese

Action: no alignment of the English *and regulations* and the Portuguese missing part of the translation \emptyset with the expressions *formularios y ventanillas* in Spanish and *formulaires et guichets* in French

2.2.2 Non-translation

The example below illustrates the non-translation of parts of the sentence. The Spanish, French and Portuguese sentences do not have part of the information existing in English, even though the Spanish misses less information than French and Portuguese.

EN – ... *and not an obligation to invest in this way*
ES – *y no [estamos hablando de] una obligación de hacerlo*
FR – *et non pas obligatoires*
PT – *e não uma obrigação \emptyset*

Action: P-block alignment of the expressions *and not an obligation* in English, *y no de una obligación* in Spanish, *et non pas obligatoires* in French, and *e não uma obrigação* in Portuguese with internal S-segment-alignments of the semantically equivalent individual words

Action: no alignment of any element of the expressions *to invest in this way* in English and *de hacerlo* in Spanish, which do not convey the same meaning, and have no correspondence in French (*obligatoires*) and Portuguese (\emptyset)

2.3 Emphatic linguistic structure

2.3.1 Tautology and pleonasm

Both tautology and pleonasm manifest redundancy, sometimes unnatural when translated into a target language. Linguistic occurrences of tautology and pleonasm are left unaligned. In the example below, the English verb *to catch up* implies that the subject *United States* is more advanced, which is already expressed implicitly in the relative clause *which is [currently] far ahead of us*.

EN – *and the European Union must make more rapid progress if it wishes to catch up, in particular with the United States, *which is currently far ahead of us*

ES – *si quiere recuperar el retraso actualmente en relación, sobre todo, con los Estados Unidos, la Unión Europea debe avanzar mucho más deprisa.*

FR – *l'Union Européenne dit avancer beaucoup plus vite, si elle veut refaire son retard dans ce domaine.*

PT – *a União Europeia tem que avançar muito mais depressa se quer recuperar o atraso actualmente existente em relação, sobretudo, aos Estados*

Unidos.

Action: no alignment of any element of the English relative *which is [currently] far ahead of us* with any element in the Romance languages

2.3.2 Repetition of words or phrases/expressions

Whenever there are repetitions of a phrase in the target sentence of a pair of aligned sentences, only the first instance of the repetition should be aligned.

EN – *very often, seaports are the first and last ports of call for cargo in Europe*

ES – *los puertos marítimos suelen ser puntos de partida y de destino de los flujos de mercancías en Europa*

FR – *les ports maritimes sont souvent les points de départ et d'arrivée des chargements en Europe*

PT – *não é raro, na Europa, os portos marítimos constituírem o ponto de partida e de chegada dos fluxos de carregamentos *na Europa*

Action: no alignment of the second Portuguese phrase *na Europa*, absent in English, Spanish and French

If a repetition in the source sentence is replaced by some anaphoric equivalent in the target sentence, no alignment takes place.

EN – *not only of our economic policy guidelines, but also of our employment policy guidelines*

FR – *dans nos lignes directrices de la politique économique, mais aussi dans celles de la politique de l'emploi*

ES – *nas nossas directrizes de política económica, sino también en las de política de empleo*

PT – *nas nossas directrizes político-económicas, mas também nas directrizes do âmbito da política do emprego*

Action: no alignment of the French and Spanish anaphora *celles de* and *las de*, whose referent in English is *guidelines*

Action: no alignment of the irrelevant Portuguese phrase *do âmbito*

2.3.3 Additional and missing information

Incomplete translations, where part of the information (even if minimal or semantically less relevant) is missing in the target sentence of the parallel sentences, should be left unaligned. Unaligned should also be any information that

is added in the target language, which did not exist in the source language. In CLUE annotations, the pieces of information that are not represented in one of the languages of the parallel sentences, are visible in the alignment matrix through blank spaces. Normally (but not always), this corresponds to irrelevant or unnecessary information.

EN – the issue of **inland ports** and intermodal terminals is very complex
ES – el tema de **los puertos de navegación marítima interior** y de las terminales intermodales es muy complejo
FR – le sujet des **ports de navigation intérieure** et des terminaux intermodaux est très complexe
PT – o tema dos ***portos marítimos e portos de navegação interior** e dos terminais intermodais é muito complexo

Action: no alignment of Portuguese *portos marítimos e*, absent in English

Action: no alignment of Spanish *marítima*, already implied in *puertos de navegación* and absent in English

EN – of large **amounts of capital** and [...] small **amounts**
ES – de los grandes capitales y [...] de los pequeños
FR – des grandes capitaux et [...] des petits
PT – dos grandes capitais e [...] dos pequenos

Action: no alignment of the English *amounts of* and *amounts*, absent in Spanish, French and Portuguese

2.4 Approximate numeric correspondence

Approximate numeric correspondences are not aligned. Only exact numeric correspondences are aligned.

EN – after **55** years of community life
ES – después de **cincuenta** años de vida comunitaria
FR – après **cinquante** ans de vie communautaire
PT – após **cinqüenta** anos de vida comunitária

Action: no alignment of the alpha numeric expressions *cincuenta* in Spanish, *cinquante* in French, and *cinqüenta* in Portuguese with the English two-digit numeric expression *55*

2.5 Mismatching pronoun and determiner

Whenever pronouns and determiners do not maintain the same form in the source and target languages altering in this way the meaning of the sentence, no alignment should take place.

EN – *the commission approved **its** proposal*
ES – *la comisión aprobó **su** propuesta*
FR – *la commission avait approuvé **votre** proposition*
PT – *a comissão [...] aprovou **a sua** proposta*

Action: no alignment of the English *its*, the Spanish *su*, and the Portuguese *a sua* with the French *votre*

2.6 Abbreviation versus full word

Abbreviations are always S-aligned (segment or block) with their full word equivalences.

EN – *i.e.*
ES – *o sea*
FR – *c'est-à-dire*
PT – *isto é*

Action: S-block-alignment of the English abbreviation with the full word equivalences in the Romance languages

2.7 Punctuation

When different languages use different punctuation marks with the same meaning, a P-alignment should be made. Spanish beginning and end of sentence question marks (*¿*) and (*?*) should be considered as indivisible tokens. Exceptionally, the beginning of sentence upside-down question mark, which is exclusive of Spanish and not of other languages, should also be S-aligned with the the end of sentence question mark.

2.7.1 Discrepancy between punctuation marks

Different but correct punctuation marks should be P-aligned.

EN – *fifthly[,] it remains unclear why mergers on the capital market should have a positive impact on the availability of risk capital*

ES – en quinto lugar[:] sigue estando oscuro por qué las fusiones en el mercado de capitales producirán efectos positivos en para la generación de capital-riesgo

FR – cinqüièmement[,] on se demande comment des fusions sur le marché des capitaux devraient avoir des effets positifs sur l’apport de capital-investissement

PT – quinto[:] o que não ficou esclarecido é por que razão é suposto as fusões no mercado de capitais terem efeitos positivos para a disponibilização de capital de risco

Action: P-alignment of the commas in English and French with the semi-colon in Spanish and Portuguese

Different mismatched (or incorrect) punctuation marks should not be aligned with any other punctuation mark in the sentence.

EN – the fewer suppliers there are, the less competition there is and the higher costs are[.]

ES – en la medida que existan menos ofertantes, será menor la competencia y los costes serán más elevados[.]

FR – moins il y a d’offrants, moins il y a de concurrence et plus les coûts sont élevés[!]

PT – quanto menos operadores, tanto menos concorrência, o que, por sua vez, se reflecte em custos mais elevados[!]

Action: no alignment of the exclamation mark in French and Portuguese with the full stop in English and Spanish

2.7.2 Optional versus obligatory punctuation

Some punctuation marks are optional and they appear in one language and not in the other. Optional punctuation non-occurring in one of the sentences of the language pair should be left unaligned. However, obligatory punctuation marks occurring in appositions (see example in Portuguese below) should be S-block-aligned with the alpha element(s) of the apposition. Conclusive conjunctions in the beginning of sentences also require a comma, such as in the Spanish sentence illustrated below.

EN – I therefore totally support parliament’s view on point 23

ES – Por consiguiente[,] apoyo totalmente la postura del parlamento sobre el punto 23

FR – J’apporte donc mon soutien inconditionnel à l’opinion de cette assemblée formulée au point 23

PT – Apoio[,] portanto[,] inteiramente[,] a opinião do parlamento sobre o nº 23

Action: P-block-alignment of the obligatory commas in the Portuguese conclusive conjunction [, *portanto*,] non-obligatory in English and French in these contexts

2.7.3 Missing and misplaced punctuation mark

Missing and misplaced punctuation marks should not be aligned with any other punctuation marks in the sentence.

EN – [;] *in other words [Ø] more risk capital allows us to use our human capital more efficiently.*

ES – [;] *en otras palabras[,] un mayor capital-riesgo nos permite utilizar de modo más eficaz nuestro capital humano.*

FR – [.] *en d'autres mots[,] plus nous disposons de capital-investissement, plus nous pouvons exploiter efficacement notre capital humain.*

PT – [;] *por outras palavras[,] mais capital de risco permite-nos utilizar mais eficientemente o nosso capital humano.*

Action: no alignment of the comma in each one of the Romance languages with the English missing comma

Action: no alignment of the full stop in French with the semi-colon in English, Spanish and Portuguese

The abbreviation *etc.* takes a left-hand comma. The comma plus the abbreviation should be S-block-aligned.

EN – *issues such as the reform of the United Nations, disarmament, arms control, the fight against Aids [etc.]*

ES – *puntos como la reforma de las Naciones Unidas, el desarme, el control de armamento, la lucha contra el Sida [y otros]*

FR – *des sujets tels que la réforme des Nations Unies, le désarmement, le contrôle de l'armement, la lutte contre le Sida [, etc.]*

PT – *pontos como a reforma das Nações Unidas, o desarmamento, o controlo de armamentos, a luta contra a Sida [, etc.]*

Action: no S-block-alignment of the abbreviation *etc.* in English with the corresponding correct comma plus abbreviation , *etc.* in French and Portuguese and with the equivalent expression in Spanish *y otros*

Action: S-block-alignment of the comma plus abbreviation *etc.* in French and Portuguese with the equivalent expression in Spanish *y otros*

Misplaced punctuation marks are typos, which should not be aligned.

EN – has also increased the need to reform the system
ES – también ha incrementado la necesidad de reformar el sistema
FR – n'a fait qu'ajouter à la nécessité de réformer le système actuel
PT – aumentou, também[.] a necessidade de reformar o sistema

Action: no alignment of the misplaced full stop in Portuguese

2.7.4 Comma + coordinating conjunction *and*

The comma plus the coordinating conjunction *and* in English correspond to the conjunction without the left-hand comma in the Romance languages.

EN – the cost of maintaining patents [, **and**] also uncoordinated tax systems
ES – los costes de mantenimiento de la patente [**y**] también la falta de coordinación entre los sistemas fiscales
FR – les coûts de maintien des brevets [**et**] les systèmes fiscaux non coordonnés
PT – os custos para a manutenção das patentes [**e**] ainda a existência de sistemas fiscais descoordenados

Action: S-block-alignment of the comma plus the coordinating conjunction in the English sentence with its equivalent conjunctions in the Romance languages (*y* in Spanish, *et* in French and *e* in Portuguese)

2.7.5 Comma + relative pronoun *which*

In English the relative pronoun *which* takes a preceding comma.

EN – it is a fortunate coincidence that the meeting is taking place in Korea soon after the historic Korean summit [, **which**] opened the way to reconciliation and future reunification
ES – pienso que ha sido una coincidencia afortunada que la cumbre se celebre en Corea poco después de la histórica cumbre coreana [**que**] abrió las puertas para la reconciliación y la futura reunificación
FR – je pense que c'est une heureuse coïncidence que le sommet se tiennne en Corée, peu après le sommet coréen historique [**qui**] a ouvert les portes de la réconciliation et de la future réunification
PT – penso tratar-se de uma feliz coincidência o facto de o encontro se realizar na Coreia pouco tempo depois da histórica cimeira coreana [**que**] abriu as portas à reconciliação e à futura reunificação do país

Action: S-block-alignment of the comma plus the relative pronoun *which* in the English sentence with the equivalent relative pronouns in the Romance

languages (*que* in Spanish and Portuguese, and *qui* in French)

3 Multiword units

Continuous multiword units are S or P-block-aligned, depending on their unambiguous or ambiguous nature. Discontinuous multiword units, where individual elements that compose the unit are placed far apart in the sentence, cannot be aligned, with the exception of multiword units with a small number of inserts. So, when the components of the discontinuous multiword unit are close to each other and a certain cohesion is maintained in the expression, even with the insertion of an external or a minimal amount of external elements, they should be aligned. Individual segments of discontinuous multiword units should not be aligned. The following subsections illustrate different types of multiword unit alignments.

3.1 Support verb constructions

In the Europarl corpus, support verb constructions are either S-aligned with semantically equivalent single verbs (many-to-one correspondence) or with other semantically equivalent support verb constructions (many-to-many correspondence). The S-alignment of the English verb *link* with the Spanish, French, and Portuguese support verb constructions *servir de enlace entre*, *s'entendent en tant que maillon entre*, and *servir de elo de ligação entre* is an example of a straightforward equivalence between a support verb construction and a single verb. Smaller block-alignments or individual segment alignments can be established in the interior of a larger block multiword expression alignment. Semantically-equivalent individual elements, independently of their part-of-speech, are P-aligned.

EN – *the commission fully approves the joint position of the council*
ES – *la comisión aprueba plenamente la posición común del consejo*
FR – *la commission approuve pleinement la position commune du conseil*
PT – *a comissão dá a sua total aprovação à posição comum do conselho*

Action: the English verbal expression *fully approves* S-block-aligns with the Spanish and French verbal expressions *aprueba plenamente* and *approuve pleinement*, and with the Portuguese support verb construction *dá a sua total aprovação a*

Action: the Portuguese adjective *total* P-aligns with the adverbs *fully* in English, *plenamente* in Spanish, and *pleinement* in French

Action: the Portuguese predicate noun *aprovação* P-aligns with the verbs *approves* in English, *aprueba* in Spanish, and *approuve* in French

The example below illustrates the equivalence between semantically related support verb constructions and a single verb in Spanish. This equivalence is annotated as an S-block alignment.

EN – we need **to draw a distinction** between north and south
ES – *debemos* **distinguir** entre norte y sur
FR – nous devons **faire une distinction** entre le nord et sud
PT – temos de **estabelecer uma diferença** entre norte e sul

Action: the English, French, and Portuguese support verb constructions *draw a distinction*, *faire une distinction*, and *estabelecer uma diferença*, S-align with the Spanish verb *distinguir*

The Europarl corpus contains alignments of support verb constructions where they are the only verbal expression possible due to lack of a suitable single verb, which is semantically equivalent to the support verb construction, in all languages. The English support verb construction *play/have role* appear in the corpus several times, as in the occurrences below.

EN – *play an increasingly predominant role*
EN – *are playing an ever-greater role*
EN – *will have a key role to play*

ES – *desempeñan un papel cada vez más predominante*
ES – *juegan un papel cada vez más fuerte*
ES – *desempeñan un papel decisivo*

FR – *jouent un rôle de plus en plus prédominant*
FR – *jouent un rôle croissant*
FR – *jouent un rôle décisif*

PT – *desempenham um papel cada vez mais predominante*
PT – *desempenham um papel cada vez maior*
PT – *desempenham um papel decisivo*

Action: S-block-alignment of all these discontinuous support verb constructions with the frequently used associated modifiers

Support verb constructions are often discontinuous. Currently, word alignment techniques, even if they contemplate multiword unit alignments, are unable to present a solution to process discontinuous expressions. For example, the English discontinuous support verb construction *bring to a conclusion*, in the example below, places the predicate noun *conclusion*, with its adnominal

modifiers, 10 words apart from the support verb *bring*. This distant placement of the different components of one unit makes the alignment of this expression impossible. The Spanish, French, and Portuguese equivalents *para que acelere la directiva sobre pensiones complementares*, *pour faire avancer la directive sur les pensions complémentaires*, and *apressar-se a apresentar a directiva relativa as pensões complementares*, adopt different stylistic variants and simpler surface structures (i.e., syntax) by transforming the support verb construction into a semantically similar verb construction, a compound verb in French and Portuguese (*faire avancer*; *apressar-se a apresentar*) and a single verb in Spanish (*acelerar*).

EN – I would urge the European Commission **to bring** the process of adopting the directive to on additional pensions **to a conclusion**

ES – insto a la comisión europea **para que acelere** la directiva sobre pensiones complementares

FR – j’insiste auprès de la commission européenne **pour faire avancer** la directive sur les pensions complémentaires

PT – exorto a comissão europeia **a apressar-se a apresentar** a directiva relativa as pensões complementares

Action: no alignment of the discontinuous English support verb construction to bring [INSERTS] to a conclusion

Action: P-block-alignment of the continuous verbal expressions between Spanish, French and Portuguese

3.2 Compounds

Compounds do not present major challenges in the alignment process and often S-align with single words in one language or another of a language pair. The examples presented below illustrate the alignment of prepositional, adverbial and verb compounds.

EN – we assisted Korea during the Asean crisis **by means of IMF loans**

ES – hayamos ayudado a Corea en la crisis de la Asean **a través de préstamos del FMI**

FR – nous ayons aidé la Corée dans la crise de l’Anase, **avec des prêts du FMI**

PT – termos ajudado a Coreia, **através de créditos do FMI**

Action: S-block-alignment of the prepositional compounds and between the English, French and Portuguese compounds, and the French single preposition *avec*

The example below shows the alignment of single adverbs with adverbial compounds.

EN – *it will certainly lead in due course to a charter that is fully **legally** binding*

EN – *sin duda alguna culminará más adelante en una carta que sea total y **jurídicamente** vinculante*

FR – *en temps voulu à l'adoption d'une charte totalement contraignante **d'un point de vue juridique***

PT – *a seu tempo conduzirá a uma carta que seja inteiramente vinculativa **em termos jurídicos***

Action: S-block-alignment of the French and Portuguese compound adverbs *d'un point de vue juridique* and *em termos jurídicos* with the one-word-adverbs in English (*legally*), and Spanish (*jurídicamente*).

The next example shows the alignment of idiomatic compound adverbial modifiers.

EN – *needs to be **crystal-clear***

ES – *deve quedar **totalmente** clara*

FR – *doit être clair **comme l'eau de roche***

PT – *deve ser **crystalinamente** clara*

Action: S-block-alignment of the entire idiomatic expression between all languages

Action: P-block-alignment of the French idiomatic compound adverb *comme l'eau de roche* with the simple adverbs in Spanish (*totalmente*), and Portuguese (*crystalinamente*)

Compound verbs, also known as complex predicates, are among compounds, those that offer a more difficult analysis through the alignment process due to their distinct and sometimes peculiar behaviour among languages. Many differences exist with regards to the grammatical structure of the verbs and their representation as verb tenses. Often two or more elements of the compound in the source language coalesce into one element in the target language, or vice versa, a new item appears in the target which is absent in the source, inserts occur in the middle of a compound in the source but not in the target or vice versa, some tenses do not exist or are rarely used in one language or the other, and so on and so forth. Often, the English compound verbs (future, past perfect, etc.) correspond to simple verbs in the Romance languages. The examples presented below illustrate some of these phenomena.

The first example shows the contrast between the English present perfect and its Romance languages equivalents, the French *passé composé*, the Portuguese

pretérito perfeito. Spanish uses the past participle form (participial adjective) as the result of the transformation of the relative clause, which occurs in English, French and Portuguese.

EN – *intergovernmental conferences we **have had** in the past*
ES – *conferencias intergubernamentales **organizadas** hasta hoy*
FR – *conférences intergouvernementales que nous **avons eues** jusqu'à présent*
PT – *conferências intergovernamentais que **tivemos** anteriormente*

Action: S or P-block-alignment of the English and French compound verbs (with auxiliary) with the simple verbs in Spanish (participial adjective) and Portuguese (*pretérito perfeito*)

EN – *the EC court **will have to apply** the articles of the charter in their judicial practices as self-binding principles*
ES – *el tribunal de justicia de la CE **tendrá que aplicar** en sus procedimientos judiciales los artículos de la carta de los derechos fundamentales como principios vinculantes*
FR – *la cour de justice **appliquera** dans sa jurisprudence les articles de la charte de droits fondamentaux à titre de principes qui l'engagent*
PT – *o tribunal da justiça da UE **irá aplicar** na sua jurisprudência os artigos da carta como princípios que o vinculam*

Action: S-block-alignment of the future tense compound in English, Spanish and Portuguese with the simple future in French

Often, adverbial inserts in the middle of compound verbs S-block-align with single verbs.

EN – *the shift in favour of more performance-oriented instruments **has also increased** the need to reform the system*
ES – *el paso hacia unos instrumentos más orientados al rendimiento **también ha incrementado** la necesidad de reformar el sistema*
FR – *le glissement observé vers des instruments plus axés sur la performance **n'a fait qu'ajouter** la nécessité de réformer le système actuel*
PT – *o facto de se estarem a privilegiar os instrumentos de maior rentabilidade **augmentou também** a necessidade de reformar o sistema*

Action: S-block-alignment of the English, Spanish, and French compound verbs (with adverbial inserts) with the Portuguese simple verb followed by the corresponding adverb

3.3 Phrasal verbs

Similarly to other multiword units, phrasal verbs need to be treated as a whole, as minimal units of meaning. However, they often are discontinuous, with the verb particle being placed quite distantly from the verb. Continuous phrasal verbs are block-aligned. Discontinuous phrasal verbs are not aligned.

EN – *the European Union should perhaps **look into** the problem*
ES – *la Unión Europea debería quizás **examinar** ese problema*
FR – *l’Union Européenne devrait peut-être **se pencher sur** ce problème*
PT – *a União Europeia deveria talvez **debruçar-se sobre** este problema*

Action: S-block-alignment of the English phrasal verb with the French and Portuguese prepositional verbs and the Spanish single verb.

3.4 Prepositional predicates

Some predicates (verbal or nominal and adjectival, e.g., support verb constructions) select specific prepositions or prepositional complements. This phenomenon has also been called as prepositional dependency, governance, or selection. Prepositional predicates may select very different types of prepositions and these prepositions may vary from language to language. Also, a predicate may select a particular preposition in one language, but select no preposition in a different language. Prepositional predicates can be considered as a specific type of multiword unit, in the sense that these predicates cannot be separated from the prepositions they select. The prepositions can be immediately attached to the predicate or they may have inserts in between. As with other types of multiword units, they can be continuous or discontinuous. The next three subsections illustrate how prepositional predicates should be aligned.

3.4.1 Prepositional verbs

Continuous prepositional verbs, i.e., transitive verbs governing prepositions, where the preposition is immediately at the right-hand side of its governing verb, are block-aligned.

EN – *our Asian partners prefer **to deal with** questions which unite us*
ES – *nuestros socios asiáticos prefieren **dedicarse a** las cuestiones que nos unen*
FR – *nos partenaires asiatiques préfèrent **s’attacher à** ce qui nous unit*
PT – *os nossos parceiros asiáticos preferem **centrar-se unicamente nas** questões comuns*

Action: P-block-alignment of the English prepositional verbs with its equivalents in the Romance languages (with an irrelevant adverbial insert in Portuguese)

The example below illustrates the alignment of a support verb construction in English with its semantically equivalent prepositional verbs in the Romance languages.

EN – *the committee on employment and social affairs* **took a special interest in** *types of supplementary pension funds*

ES – *la comisión de empleo y de asuntos sociales* **se ha ocupado en especial de** *las modalidades de la asistencia suplementaria a la tercera edade*

FR – *la commission de l’emploi et des affaires sociales* **s’est en particulier occupée des** *différentes formes de retraite complémentaire*

PT – *a comissão do emprego e dos assuntos sociais* **debruçou-se em especial sobre** *as possibilidades existentes para regimes complementares de reforma*

Action: S-block-alignment of the English support verb construction with the its semantically equivalent prepositional verbs in the Romance languages

Discontinuous prepositional verbs, where the preposition is distant from the verb, are not aligned. Discontinuous prepositional verbs may have complex argument structures. The next example illustrates prepositional verbs with the following argument structures: [*congratulate N1 on N2*] in English, [*felicitar a N1 por N2*] in Spanish (where the verb *felicitar* governs the preposition *a*), [*féliciter N1 pour N2*] in French, and [*felicitar N1 por N2*] in Portuguese. It is unfeasible to annotate discontinuous preposition-dependency.

EN – *I too should like* **to congratulate** [*NE*] **on** *his excellent report*

ES – *también yo quisiera* **felicitar a** *mi colega* [*NE*] **por** *su excelente informe*

FR – *je voudrais* **féliciter** *moi aussi mon collègue* [*NE*] **pour** *son excellent rapport*

PT – *também eu gostaria de* **felicitar** *o meu colega* [*NE*] **pelo** *seu excelente relatório*

Action: S-segment-alignment of the individual words of the prepositional verb between the different languages

3.4.2 Prepositional nouns

Prepositional nouns are predicate nouns that select a given complement introduced by a particular preposition. Continuous prepositional nouns, i.e., nouns governing prepositions, where the preposition is immediately at the right-hand side of its governing predicate noun, are block-aligned.

EN – as the demographic changes increase **the need for** retirement products
ES – a medida que los cambios demográficos han aumentado **la necesidad de** productos de jubilación
FR – poussés par les évolutions démographiques et **la demande de** produits financiers liés aux retraites
PT – dado que as alterações demográficas aumentam **a necessidade de** "produtos de reforma"

Action: S-block-alignment of the prepositional nouns among all language combinations

3.4.3 Prepositional adjectives

Prepositional adjectives are predicate adjectives that require an auxiliary or support verb and select a given complement introduced by a particular preposition. Prepositional adjectives can be continuous or discontinuous. Continuous prepositional adjectives are S-block-aligned, but discontinuous ones are not aligned. The example below illustrates the alignment of continuous prepositional adjectives, namely the prepositional adjective *informed about* (with the exception of the Spanish sentence, whose prepositional adjective contains an adverbial insert).

EN – calling on the commission to keep us regularly **informed about** recent developments
ES – pidiendo a la comisión que nos mantenga **informados** periódicamente de lo que vaya ocurriendo
FR – appelant la commission à nous tenir régulièrement **informés des** derniers développements de ce dossier
PT – em que se apela à comissão para que nos mantenha regularmente **informados acerca dos** progressos que se forem realizando

Action: S-block-alignment of the prepositional adjectives *informed about* in English, *informés des* in French, and *informados acerca dos* in Portuguese

Action: no alignment of the Spanish prepositional adjective, because it contains the adverbial insert *periódicamente* between the adjective *informados* and the associated preposition *de*

Prepositional adjectives can be part of larger expressions with specific meanings, such as support verb constructions. These support verb constructions can be paraphrased into verbs or other types of support verb constructions. The example below shows prepositional adjectives in English and Spanish occurring in support verb constructions (*to be aware of* and *ser consciente de*), which were transformed into the single verb *connaître* in French and into the support verb construction *ter consciência de* in Portuguese. This support verb construction

in Portuguese uses a predicate noun that is semantically equivalent to the source predicate adjective (*aware* (predicate ADJ) - *consciente* (predicate ADJ) - *consciência* (predicate N)).

EN – *we are all aware of the problems involved*
ES – *todos somos conscientes de la problemática*
FR – *nous connaissons la problématique*
PT – *temos todos consciência da problemática*

Action: S-block-alignment of the English and Spanish prepositional adjectives (internal alignments of a larger block)

Action: S-block-alignment of the expressions *we are all aware of* in English, *todos somos conscientes de* in Spanish, *nous connaissons* in French, and *temos todos consciência da* in Portuguese

3.5 Named entities

English named entities, such as *European Parliament*, or *European Investment Bank* S-align with the corresponding named entities in Spanish (*Parlamento Europeo*, *banco europeo de inversiones*), French (*Parlement Européen*, *banque européenne d'investissements*), and Portuguese (*Parlamento Europeu*, *banco europeu de investimento*). Named entities in the Romance languages have a distinct internal structure, different from those in English. However, usually named entities are translated as a whole in all contexts, and they should be S-block-aligned.

EN – *committee on economic and monetary affairs*
ES – *comisión de asuntos económicos y monetarios*
FR – *commission économique et monétaire*
PT – *comissão dos assuntos económicos e monetários*

Action: S-block-alignment of the entire named entity in all languages

3.6 Date and time expressions

Date and time expressions are normally continuous and they are S-block-aligned between one language and another.

EN – *the commission approved its proposal on 20 September*
ES – *la comisión aprobó su propuesta el día 20 de Septiembre*
FR – *la commission a approuvé sa proposition le 20 de Septembre*
PT – *a comissão aprovou a sua proposta em 20 de Setembro*

Action: S-block-alignment of the date expression with internal S-block-alignment of the dates *20 September* in English, *20 de Septiembre* in Spanish, *20 de Septembre* in French, and *20 de Setembro* in Portuguese

EN – *the vote will take place [tomorrow at **12:30 p.m.**]*
ES – *la votación tendrá lugar [mañana a las **12.30 horas**]*
FR – *le vote aura lieu [demain à **12h30**]*
PT – *a votação terá lugar [amanhã, às **12h30**]*

Action: S-block-alignment of the time expression with internal S-block-alignment of the time *12:30 p.m.* in English, *12.30 horas* in Spanish, and *12h30* in French and Portuguese

3.7 Lexical bundles

English lexical bundles S-align with the corresponding lexical bundles in the Romance languages.

EN – *I believe that*
EN – *I should like that*

ES – *yo creo que*

FR – *je crois que*

PT – *creio/acredito que*
PT – *gostaria de*
PT – *não posso deixar de*

Action: S-block-alignment of the lexical bundle among all languages

3.8 Idiomatic expressions

Idiomatic expressions block-align with current non-idiomatic words or other expressions (idiomatic or non-idiomatic).

EN – *which **began** four years ago*
ES – *que **empezó** hace cuatro años*
FR – *qui **a vu le jour** il y a quatre ans*
PT – *que **se iniciou** há quatro anos*

Action: P-block-alignment of the non-idiomatic verbs in English, Spanish and Portuguese with the idiomatic verbal expression in French

EN – which produce opposing **reactions**
FR – qui suscitent des **états d’âme** divergents
ES – que suscitan dos **estados de ánimo** encontrados
PT – que suscitam dois **estados de espírito** divergentes

Action: P-block-alignment of the non-idiomatic English noun with the idiomatic compound noun in the Romance languages

3.9 Domain terms

Domain terms S-block-align with equivalent terms in other languages. Domain terms can be embedded, therefore, internal alignments can be made inside a domain term.

EN – efficient **venture capital markets** are also a precondition for the development of entrepreneurship in Europe
ES – unos **mercados eficientes de capital-riesgo** constituyen también una condición previa para el desarrollo del espíritu de empresa en Europa
FR – des **marchés de capital-investissement** son également une condition sine qua non au développement de l’esprit d’entreprise en Europe
PT – para o desenvolvimento do espírito empresarial na Europa, é também necessário, como condição prévia, a existência de **mercados de capital de risco eficientes**

Action: the English domain term *venture capital markets* S-aligns with the equivalent terms *marchés de capital-investissement* in French, and *mercados de capital de risco* in Portuguese, with the internal alignment for the term *venture capital*, corresponding to *capital-investissement* in French, and *capital de risco* in Portuguese. The Spanish equivalent contains the adjectival insert *eficientes*, which represents an obstacle to the alignment of the domain term.

3.10 Other expressions

Other expressions corresponding to translation units, are also S-block-aligned.

EN – I would like to emphasise once again that
EN – quiero señalar de nuevo que
FR – j’attire une fois de plus votre attention sur le fait que
PT – chamo novamente a atenção para o facto de

Action: S-block-alignment of the entire expression among all languages

4 Lexical versus non-lexical realization

4.1 Determiners and zero determiners

Most challenges to the alignment of determiners consist on the alignment of the definite articles. The presence of zero articles is very common in English, and less frequent in the Romance languages. Aligning a zero determiner with a lexically realized determiner implies association to the noun. Determiners are only aligned together with the noun (single or compound) when they do not appear in one of the languages of an alignment pair. When determiners appear in both languages of the alignment pair, they are segment aligned, i.e., they are aligned individually.

EN – *it was the critical moment when our exports to **Asia** exceeded our exports to the United States*

ES – *el momento crítico en que nuestras exportaciones hacia **Asia** superaron las dirigidas a los Estados Unidos*

FR – *ce fut l'année où nos exportations vers **l'Asie** ont dépassé celles vers les États-Unis*

PT – *o momento crítico em que as nossas exportações para **a Ásia** excederam as destinadas aos Estados Unidos*

Action: P-block-alignment of [Determiner + N] with [Determiner-Ø + N]

The alignment of compounds with determiners with compounds with no determiners implies that the lexically realized determiner is associated to the compound. However, an internal alignment is possible that maps the head of the compound with no determiner with the head of the compound with a determiner whenever the word order is regular, i.e., [Determiner + head Noun + Modifier(s)].

EN – *(through) **Ø** indirect taxes*

ES – *(mediante) **los** impuestos indirectos*

FR – *(par le biais) **des** impôts indirects*

PT – *(através d)**os** impostos indirectos*

Action: P-alignment of the compound with the compound modified by the determiner

Action: P-alignment of the head of the compound having a determiner with the head of the compound having no determiner

In case of a contraction (such as in the case of the Portuguese sentence in the example above), the alignment of the determiner is established in the contraction, and no alignment of the determiner with the noun takes place. The following example illustrates the segment alignment of [Preposition + Determiner]

in French and Portuguese with single prepositions in English and Spanish, in which languages the proper name does not take a preceding definite article. This is a pragmatic solution, not a linguistic motivation, as we will demonstrate in Section 5.2.

EN – **in** \emptyset *Japan*
ES – **en** \emptyset *Japón*
FR – **au** *Japon*
PT – **no** *Japão*

Action: segment alignment of a zero determiner in English and Spanish with a determiner in a contraction (i.e., attached to a preposition) in French and Portuguese

Possessive and some relative pronouns also require a determiner in Portuguese (and sometimes, also in Spanish). When contrasting English and French possessive pronouns, one word in the these languages corresponds to two words in Portuguese (and sometimes, in Spanish).

EN – *yet this concerns the majority of the population, as it is **their** money that the European institutions count on using in order to fund the measures to facilitate private investment*

ES – *sin embargo la mayoría de la población está afectada puesto que es con **su** dinero que las instituciones europeas piensan financiar las medidas para facilitar inversiones privadas*

FR – *la majorité de la population est cependant concernée car c’est avec **son** argent que les institutions européennes comptent financer les mesures facilitant les investissements privés*

PT – *a maioria da população é no entanto envolvida, pois é com **o seu** dinheiro que as instituições europeias contam financiar as medidas destinadas a facilitar os investimentos privados*

Action: S-block-alignment of the Portuguese two-word pronoun with the single word pronouns in English, Spanish, and French

4.2 Pronoun-dropping phenomena

4.2.1 Subject pronoun drop

The contrast of a language with overt pronouns with a language with omitted or null pronouns is a recurring phenomenon in the alignment process (e.g., *I went* = \emptyset *fui*). Our guidelines establish that a P-alignment with a verb should be made if a personal pronoun is only available in one language of the parallel sentence pair. If the personal pronouns are overt in both languages, an S-alignment should be made between these pronouns as single segments, and a

separate alignment should be made for the verbs.

EN – *which is something we regret*
ES – *y lamentamos esta situación*
FR – *ce que nous ne pouvons que regretter*
PT – *facto que lamentamos*

Action: S-block-alignment of the English [Pronoun + V] with the Spanish and Portuguese [Pronoun- \emptyset + V]

Action: S-segment-alignment between the French and English lexically realized 1st person plural pronouns

Action: P-block-alignment of the French overt pronoun focus construction with its semantically [pronoun or zero-pronoun + V] equivalents in other languages

4.2.2 Empty relative pronoun

English allows empty relative pronouns, which need to be lexically realized in the Romance languages. For example, the English relative [N *that we have had*] is equivalent to the implicit relative [N *we have had*], in the example below. The relative pronoun is obligatory in the Romance languages (French and Portuguese), unless the relative is transformed into a participial adjective (i.e., a paraphrase of the relative), such as is the case of the Spanish example illustrated below.

EN – *intergovernmental conferences \emptyset we have had in the past*
ES – *conferencias intergubernamentales organizadas hasta hoy*
FR – *conférences intergouvernementales que nous avons eues jusqu'à présent*
PT – *conferências intergovernamentais que tivemos anteriormente*

Action: S-block-alignment of the English empty pronoun relative with the relatives with realized pronouns in French and Portuguese

Action: P-block-alignment of the relatives in English, French and Portuguese with the Spanish paraphrasial semantically stronger participial adjective

4.3 Relatives versus participial adjectives

Often, relative clauses are paraphrases of participial adjectives. The Europarl corpus test set demonstrates that translators often adopt a shorter version of

a relative construction, where the English sentence contains a relative and the Romance languages contain participial adjectives with identical meaning to the relative, in the same context. This is what is shown in the example below.

EN – *policy* **which is driven** *by profit and that* **which is driven** *by environmental considerations*

ES – *política* **centrada** *en el beneficio y política* **centrada** *en el medio ambiente*

FR – *entre une politique* **orientée** *vers les bénéfiques et une politique* **orientée** *vers l’environnement*

PT – *uma política* **orientada** *para o lucro e uma política* **orientada** *para o ambiente*

Action: P-block-alignment of the English relative with the participial adjective in the Romance languages

5 Other linguistic phenomena

5.1 Free noun adjuncts

When English noun adjunct phrases correspond to compounds, they are S-block-aligned with their equivalents in each individual Romance language. For example, the English compound, which also corresponds to the terminology domain, *European Investment bank* ([Adj N N]) S-block-aligns with its Portuguese equivalent *banco europeu de investimento* ([N Adj [de N]]). When these adjunct phrases do not correspond to compounds (i.e., when they appear in free noun phrases), they S-align individually. The S-alignment of English non-compound (free) noun adjuncts with their semantic equivalences in the Romance languages is illustrated in the example below. The English linguistic structures represented are [N N] and [N N N], which translate into the linguistic structures [N [de N [de N]]] and [N [de N]] in the Romance languages, where *de* is the preposition *of*. For example, in the second sequence of free noun adjuncts in English, the noun *mechanisms* S-aligns with *mecanismos*, *mécanismes*, and *mecanismos*, and *codetermination* with *de cogestión*, *de codécision*, and *de co-decisão*, in Spanish, French and Portuguese, respectively.

EN – *it* **employee participation systems** *linked with* **codetermination mechanisms** *therefore need to be in place*

ES – *por esta razón es preciso que existan también* **sistemas de participación de los trabajadores**, *que estén vinculados a* **mecanismos de cogestión**

FR – *c’est pourquoi il faut aussi* **des systèmes de participation des travailleurs**, *liés à des* **mécanismes de codécision**

PT – *it por este motivo, também é necessário que existam sistemas de participação dos trabalhadores, aliados a mecanismos de co-decisão*

Action: S-segment-alignment of individual elements of the English phrases (free noun adjuncts) *employee participation systems* and codetermination mechanisms

Action: P-block-alignment of [de N] in the Romance languages with the second and third position nouns in English

5.2 Contracted forms

In NLP tasks, contractions are problematic for several reasons: (i) two or more words with different parts-of-speech overlap, which makes syntactic analysis and generation difficult; (ii) in cross-language analysis, the contrast between languages that have contractions and languages that do not have them, or do not have them in the same contexts, presents additional difficulties. The alignment of one segment that corresponds to a contracted form in one language with the corresponding segments where elements are not contracted in the other language of the parallel pair is pragmatically (not linguistically) motivated. S-alignments should be used in those cases when a contraction is explicitly translated in the other language, either as a contraction (in between Romance languages) or as separate words translating the different elements of the contraction. P-alignments with the head element of the contraction should be used if none of the elements (words) of the contraction appear in the contrasting language. If only one element (word) of the contraction appears in the contrasting language, an S-alignment should be used for the overt element.

EN – **in that** *sense*
ES – **en ese** *sentido*
FR – **en ce** *sens*
PT – **nesse** *sentido*

Action: S-block-alignment of the lexical bundle *in that sense* with the full equivalents in the Romance languages

Action: P-block-alignment of the Portuguese contraction of the preposition and the demonstrative pronoun *nesse* with the two individual segments in the Romance languages

5.3 Singular versus plural (related to determiner)

Often translation does not preserve number inflectional features. However, because most often these morphological features do not result into any substantial

semantic change, or no semantic change at all, the decision of S-aligning nouns with different number seems justified. From a translation point-of-view, discrepancies in number are often justified by a variety of factors. However, the general translation principle of maintaining closeness to the original makes the most sense, unless there is a language parameter that forces a structure change or some good stylistic reason for changing it.

EN – *which at the same time promotes* **innovation**
ES – *que, por un lado, fomenta* **las innovaciones**
FR – *qui encourage d’une part* **les innovations**
PT – *que, por um lado, promove* **inovações**

Action: S-segment-alignment of number morphological differences between English (singular) and the Romance languages (plural)

Action: S-block-alignment of overt determiner plus noun in Spanish and French with the noun (with zero determiner) in English and Portuguese

Some determiners, such as *every* and *each* in English or *cada* in Portuguese are always used with a singular noun form. This is a grammatical constraint, not a semantic one. From the semantic point of view, these determiners have a plural referent. The nominal constructions where they occur, S-block-align with constructions occurring with determiners that take plural nouns, such as the Spanish *todos* or the French *toutes*.

EN – *in* **every official language** *of the union*
ES – *en* **todos los idiomas oficiales** *de la unión*
FR – *dans* **toutes les langues officielles** *de l’union*
PT – *em* **cada uma das línguas oficiais** *da união*

Action: S-block-alignment of the full expression in bold (grammatical singular and plural)

Action: internal S-segment-alignment

5.4 Active versus passive

In the case of a language contrast between the active and the passive voice, all elements that are translations of each other should be aligned together. If some elements are not present in the translation, such as a missing or null subject, in one of the languages of the alignment pair, no alignment should take place.

EN – **before** *new member states* **are admitted**
ES – **antes de la incorporación de** *nuevos miembros*
FR – **avant l’admission de** *nouveaux membres*

PT – **antes da entrada de** *novos membros*

Action: P-block-alignment of the passive verb in English with the semantically equivalent [predicate nouns + preposition] in the Romance languages

Action: S-alignment (segment or block) of the prepositions *before* in English, *antes de* in Spanish, *avant* in French, and *antes da* (preposition + determiner contraction) in Portuguese

Often passive constructions, where the agent is omitted, are semantically equivalent to impersonal constructions. These contrasts appear frequently in the Europarl corpus and were P-block-aligned.

EN – *the charter can be monitored from a legal perspective*

ES – *la carta puede controlarse jurídicamente*

FR – *la charte pourrait faire l'objet d'un contrôle juridique similaire*

PT – **é possível controlar** *jurídicamente a carta*

Action: P-block-alignment of the English passive construction with the active constructions in Spanish, French and Portuguese

The passive future is S-block-aligned with the active future, as illustrated in the example below.

EN – *which we will send [...] in which we will analyse*

ES – *que enviaremos [...] en la que analizaremos*

FR – *qui sera envoyée [...] dans laquelle seront analysés*

EN – *que remeterá [...] na qual se analisarão*

Action: S-block-alignment of the French passive future with the active future in English, Spanish and Portuguese

5.5 Coordination

It occurs with a certain frequency that, in coordinated noun phrases, one language repeats the coordinator and another language does not repeat it. For example, in the English noun phrase coordination *linked to poverty and insecurity*, the preposition *to* is used only in the first conjoin noun phrase, and not in the second one. In Portuguese, in the noun phrase coordination *associado a pobreza e a insegurança*, the preposition is repeated in both noun phrases. The example below shows the contrast between English and the Romance languages with regards to the lexical realization or non-realization of the preposition in the second conjoin.

EN – *which we will send to the council and Ø parliament*

ES – que enviaremos **al** consejo **y al** parlamento
FR – qui sera envoyée **au** conseil **et au** parlement
PT – que remeterá **ao** conselho **e ao** parlamento

Action: P-block-alignment of the second conjoin in English with the second conjoin in the Romance languages

5.6 Noun pre-modifiers

There are occurrences of variance in the Europarl corpus with regards to noun pre-modifiers. That reflects more on the translator’s choice in changing the original words that appear in the source language than on structural differences between languages. Since often determiners and quantifiers do not have a strong meaning associated to them, they are often not literally translated. When contrasting languages, there is often a mismatch on these linguistic elements, without any significant change in meaning. These contrasts on the use of different pre-modifiers are annotated as S-alignments.

EN – even in **those** countries where it does operate it operates in a different way from the system across the Atlantic

ES – e incluso en **aquellos** países en los que realmente funciona, lo hace de un modo diferente al del sistema del otro lado del atlántico

FR – et même dans **les** pays où le principe du capital-investissement fonctionne, il le fait selon des modalités différentes de ce que l’on trouve outre-Atlantique

PT – mesmo **nos** países em que funciona, fá-lo duma maneira diferente do sistema que existe do outro lado do Atlântico

Action: P-segment-alignment of the noun pre-modifiers among all language combinations

EN – we should use **that** public funding for **those types of** project which are most difficult to finance through the private sector

ES – deberíamos utilizar **esa** financiación pública para **aquel tipo de** proyectos que tienen mayor dificultad para ser financiados por el sector privado

FR – nous devrions recourir **au** financement public pour **les** projets que le secteur privé boude

PT – **o** financiamento público deveria ser utilizado para **os** projectos que registam maiores dificuldades em serem financiados pelo sector privado

Action: P-segment-alignment of the first occurrence of noun pre-modifiers among all language combinations

Action: segment or block P-alignment of the second occurrence of noun pre-modifiers

5.7 Anaphoric reference

In alignment tasks, anaphora presents difficult challenges. In the source language there might be no anaphora, but there is anaphora in the target language. In the example below, French, and Portuguese sentences contain a pronoun and an anaphoric pronoun + common noun as co-referents of the proper name *Korea*. The anaphora is annotated as a P-alignment.

EN – *it is not acceptable that we assisted **Korea** during the Asean crisis by means of IMF loans and suchlike, only for **Korea** still to be subsidising its shipyards*

ES – *no resulta procedente que hayamos ayudado a **Corea** en la crisis de la Asean a través de préstamos del FMI, etc. y que **Corea** siga subvencionando sus astilleros*

FR – *il n'est pas acceptable que nous ayons aidé la **Corée** dans la crise de l'Anase, avec des prêts du FMI, etc. et qu'**elle** continue à subventionner ses chantiers navals*

PT – *é inadmissível que, depois de termos ajudado a **Coreia**, através de créditos do FMI, etc., na crise da ASEAN, **este país** continue a subvencionar agora os seus estaleiros navais*

Action: P-alignment (segment or block) of the named entity (place name) in English *Korea* and Spanish *Corea* with the anaphoric pronoun *elle* in French and the deictic noun phrase *este país* in Portuguese

5.8 Antonyms and negation constructions

Often linguistic strategies, such as the negation of the opposite, are used as paraphrases or ways of saying the same thing using different wording. These contrasts should be S-aligned.

EN – *the countries of Asia **have not** unfortunately been in favour of that proposal*

ES – *los países de Asia **desgraciadamente no han sido** favorables a dicha propuesta*

FR – *les pays d'Asie **ont malheureusement rejeté** cette proposition*

PT – *os países da Ásia, **infelizmente, não se mostraram** favoráveis a esta proposta*

Action: S-block-alignment of the negation together with the adverbial insert (in English and French)

EN – *things are not going well*

ES – *las cosas* **van mal**
FR – *les choses* **vont mal**
PT – *as coisas* **estão a correr mal**

Action: S-block-alignment

EN – **surely** *the Moluccas* **cannot be overlooked**
ES – **no puede ser que no se trate el tema de** *las Molucas*
FR – **il est impossible que** *le sort des moluques* **ne soit pas abordé**
PT – **não é possível que** *as molucas* **fiquem por discutir**

Action: no alignment of the English with the Romance languages due to the discontinuity

5.9 Flexible/loose paraphrasing constructions

Some contrasts between languages are looser, i.e., they correspond to flexible paraphrases where the target languages contain less literal translations, less attached to the source language structure.

EN – *and we shall vote against it*
ES – *y merece nuestra condena*
FR – *et dénonçons*
PT – *e merece a nossa condenação*

EN – **1993 was a significant year**
ES – *el año 1993 es una fecha notable*
FR – *l'année 1993 est à marquer d'une pierre blanche*
PT – *1993 é uma data charneira*

EN – **the question is** *now* **no longer whether**
ES – *hoy no sólo se trata de si*
FR – **la question** *d'aujourd'hui* **n'est pas de savoir si**
PT – **o que está** *hoje* **em debate não é apenas saber se**

Action: S-block-alignment with the adverbial insert

5.10 Different parts-of-speech with same semantics

5.10.1 Verb versus noun predicates (process nouns)

Verb predicates S-align with corresponding nominal predicates (process nouns).

EN – we must use all the financial instruments at our disposal **to rapidly develop** the market

EN – es preciso utilizar todos los instrumentos financieros disponibles **para un rápido desarrollo ulterior** del mercado

FR – il faut utiliser tous les instruments financiers disponibles **pour développer rapidement** le marché

PT – todos os instrumentos financeiros disponíveis deverão ser aplicados **para continuar a desenvolver rapidamente** o mercado

Action: S-block-alignment (with internal segment P-alignments)

Action: no segment alignment of the Portuguese aspectual *continuar a*

5.10.2 [of ADJ N] versus [ADV ADJ]

Constructions such as [of ADJ N] S-align with constructions of the type [ADV ADJ]

EN – which are **of particular importance** in the region

ES – **de particular importancia** en la región

FR – **particulièrement importants** dans la région

PT – **que revestem uma especial importância** na região

Action: S-block-alignment

Action: segment P-alignment of the internal semantically equivalent words with different part-of-speech

5.10.3 [ADJ N] versus [ADJ Prep N]

Adjective plus noun constructions S-align with adjective plus preposition plus noun constructions.

EN – on the **prudential** rules

ES – sobre las **reglas de prudencia**

FR – sur les **règles prudentielles**

PT – sobre as **normas prudenciais**

Action: S-block-alignment

5.10.4 Gerundive and process nouns

English gerundive (-ing forms) give rise to a range of different translations

EN – between 80 and 90% of funds raised from venture capital normally go towards **hiring** more people

ES – entre el 80 y el 90% de la financiación del capital-riesgo se destina normalmente a **contratar** a más personal

FR – entre 80 et 90% des fonds qu’il mobilise sont normalment consacrés à l’**embauche**

PT – entre 80 e 90% dos fundos angariados pelos mercados de capital de risco são, normalmente, canalizados para o **recrutamento de mais pessoal**

Action: P-block-alignment

5.10.5 [to-V + N] versus [(Prep +) N Prep N]

English to-infinitive plus a noun block-align with constructions of the type [(Prep +) N Prep N] in the Romance languages.

EN – we have two action plans, one **to aid** commerce and the other **to promote** investments

ES – tenemos en marcha dos “action plans”, uno para **facilitar** el comercio y otro **para la promoción de** inversiones

FR – je rappelle que deux “action plans” sont en chantier: le premier, destiné à **faciliter** le commerce et le second, visant à **encourager** les investissements

PT – foram iniciados dois planos de acção, um visando a **promoção das** trocas comerciais e outro destinado à **promoção de** investimentos

EN – the plan is **to favour** pension plans

ES – está previsto **favorecer** los fondos de pensiones

FR – on prévoit de **favoriser** des fonds de pension

PT – prevê-se o **favorecimento** dos fundos de pensão

EN – initiatives will be taken **to combat** this scourge of our times

ES – se tomarán iniciativas **para luchar contra** esta plaga de nuestro tiempo

FR – mais des initiatives **en vue d’enrayer** ce fléau de notre temps

PT – sendo ainda tomadas iniciativas **de combate a** este flagelo do nosso tempo

EN – this is not a **desire to exclude** anything

ES – sino simplemente el **deseo de utilizar**

FR – il ne s’agit pas là d’**une mesure d’exclusion**

PT – não se trata de **um desejo de exclusão**

Action: S-block-alignment

5.11 Impersonal constructions

Impersonal relative constructions block-align with participial adjectives.

EN – **we must fully support** *the demands that have been made*
ES – **hay que apoyar plenamente** *las exigencias que se han formulado*
FR – **il faut par conséquent appuyer** *les requêtes formulées*
PT – *as reivindicações formuladas* **deverão ser plenamente apoiadas**

Action: P-block-alignment of the English *we must*, the Spanish *hay que*, and the French *il faut*

Action: internal segment S-alignment of the English, Spanish and French adverbs and verbs, and internal segment P-alignment of the Portuguese verb

5.12 Romance languages double negation (+ coordination)

Romance languages have double negation, which contrast with English, which does not allow such type of constructions.

EN – *it is not, therefore, surprising that* **there is**, *in this context*, **no** *real integration* **or** *genuine political dialogue*
ES – *no es nada sorprendente, entonces, que en ese contexto*, **no haya ni verdadera integración ni** *verdadero diálogo político*
FR – *rien d'étonnant donc, qu'il n'y ait dans ce contexte*, **ni** *intégration véritable, ni dialogue politique véritable*
PT – *assim, não é de espantar que, nesse contexto*, **não exista verdadeira integração nem** *verdadeiro diálogo político*

Action: P-block-alignment of the negative relative existential with adverbial (insert) *that there is, in this context, no* in English, *que en esse contexto, no haya* in Spanish, *qu'il n'y ait dans ce contexte* in French, and *que, nesse contexto, não exista* in Portuguese.

Action: segment P-alignment of negation and negation connector *no* – *or* in English, *ni* – *ni* in Spanish, *n'* – *ni* in French, and *Ø* – *nem* in Portuguese

6 Idiosyncrasies of languages

6.1 Portuguese inflected infinitive (peculiar verb tense)

The Portuguese inflected infinitive has no correspondence in other languages. So, when translating from Portuguese, the translator is forced to use different

verbal constructions in the target language.

EN – **the possibility that** *the commission may have a certain margin for discretion*

ES – **la posibilidad de que** *la comisión tenga un cierto margen de discrecionalidad en cuanto a la concesión de ...*

FR – **la possibilité que** *la commission jouisse d'une certaine liberté quant à la concession de périodes transitoires*

PT – **a possibilidade de** *a comissão ter uma certa margem de discricionalidade no que diz respeito à...*

Action: P-alignment of the whole construction with the N subject

EN – *we are therefore perfectly able to accept that the commission should commit itself to granting these requests provided that the member states involved can objectively justify that it is difficult for them to immediately apply the provisions of the regulation*

ES – **podemos aceptar perfectamente que** *la comisión se comprometa a acceder a esas demandas a condición de que los estados miembros afectados justifiquen de forma objetiva la dificultad con la que pueden encontrarse para aplicar de forma inmediata las disposiciones del reglamento*

FR – **nous pouvons parfaitement concevoir que** *la commission s'engage à accéder à ces demandes pour autant que les états membres concernés justifient, de façon objective, des difficultés qu'ils rencontrent dans l'application immédiate des dispositions du règlement*

PT – **a comissão pode assumir o compromisso de deferir** *essas solicitações sob condição de os estados-membros requererem de forma objetiva as dificuldades que se antepõem a uma aplicação imediata das disposições do regulamento*

Action: P-alignment

EN – *our partners have accepted this principle in return for a guarantee that wisdom and common sense will prevail when it comes to selecting the issues for discussion*

FR – **nos partenaires ont accepté ce principe étant entendu que** *la sagesse et le bon sens prévaudront dans le choix des sujets de discussion*

ES – **nuestros socios han aceptado este principio a cambio de la garantía de que** *la sabiduría y el sentido común prevalecerán al seleccionar los temas de discusión*

PT – **os nossos parceiros aceitaram este princípio sob condição de** *a sabedoria e o senso comum prevalecerem na escolha dos temas de debate*

Action: P-alignment

6.2 English infinitive (*to + V*)

The English infinitive S or P-block-aligns with the single verbs in the Romance languages.

EN – *I would like **to thank** all the speakers for their contribution to the debate*

ES – *quiero **dar las gracias a** todos los oradores por sus aportaciones al debate*

FR – *je tiens, tout d’abord, à **remercier** tous les orateurs pur leur contribution au débat*

PT – *desejo **agradecer a** todos os oradores as suas contribuições para o debate*

Action: S-block-alignment of *to + infinitive* with the infinitive (+ preposition, if applicable) in the Romance languages

6.3 French negation (*ne pas*)

French negation can be considered some type of discontinuous multiword unit, because it takes the verb in between *ne* and *pas*.

EN – *but the availability of capital is **not** the only threshold*

ES – *sin embargo, la disponibilidad de capital **no** representa el único obstáculo*

FR – *mais la disponibilité de capital **n’est pas** la seule entrave*

PT – *todavia, a disponibilidade de capital **não** constitui a única barreira*

Action: no segment-alignment of the French negation with the negation in English, Spanish and Portuguese.

Action: P-block-alignment of the French verb plus negation *n’est pas* with its equivalents in English, Spanish and Portuguese.

EN – *it **cannot be said often enough** that the European Union involves pooling our strenghts*

ES – *en la Unión Europea **no se podrá insistir con suficiente frecuencia** en que es necesario aunar las fuerzas*

FR – *dans l’Union Européenne, **on ne dit jamais assez** que l’union fait la force*

PT – *na União Europeia, **nunca é demais** chamar a atenção para a importância da congregação de forças*

Action: P-block-alignment of the predicate plus negation

6.4 English apostrophe

The English possessive apostrophe does not exist in the Romance languages. It normally corresponds to the preposition *de* (*of*) in the Romance languages.

EN – a policy that ignores economic realities and people’s needs
ES – una política que ignore las realidades económicas y las necesidades **de** la población
FR – une politique qui ignore les réalités économiques et les besoins **des** populations
PT – uma política que ignora as realidades económicas e as necessidades **das** populações

Action: P-segment-alignment of English apostrophe ’s with [PREP] *de* in Spanish or the contraction [PREP + DET] *des* in French and *das* in Portuguese

6.5 Focus constructions

Focus is a linguistic phenomenon which is used to emphasize a constituent in a sentence. Focus constructions result from the use of semantico-syntactic mechanisms that position a particular element in prominence with regards to the other elements in a sentence. Depending on the language, certain focus expressions or emphatic markers have become frozen or semi-frozen. Focus constructions P-block-align with their non-focus equivalents.

EN – satisfaction **at** the convention and bitterness **at** the outcome of the referendum
ES – la satisfacción **por** la convención y la tristeza **por** el resultado
FR – la satisfaction **pour ce qui est de** la convention et l’amerture **quant aux** résultats du référendum
PT – a satisfação **pela** convenção e o pesar **face ao** resultado

Action: P-block-alignment of the French focus construction (redundant linguistic data *ce qui est de*) with its non-focussed equivalents in English (*at*), Spanish (*por*), and Portuguese (*pela*)

Action: P block or segment-alignment of the prepositions or prepositional compounds *at* in English, *por* in Spanish, *quant aux* in French, and *face ao* in Portuguese

EN – **which is something** we regret
ES – y lamentamos **esta situación**
FR – **ce que nous ne pouvons que** regretter
PT – **facto que** lamentamos

Action: S-block-alignment of the English, French and Portuguese focus constructions *which is something Ø* and *ce que nous ne pouvons que* and *facto que* with their non-focussed equivalent *esta situación* in Spanish

6.6 Sociolinguistic differences: register and forms of courtesy

The Europarl corpus is a corpus of European Parliament speeches, i.e., the language is used in a particular social setting, where distinct languages use different registers. For example, the English *you* corresponds, in some instances, to the more formal and impersonal style of addressing people, in this case, members of the parliament, in Portuguese *senhores deputados*. Within this register, the English pronoun *you* S-block-aligns with these other more formal Romance expressions.

6.6.1 Addressing

Different languages have different form of addressing people and different forms of courtesy.

EN – *there should be none of the problems **you** have mentioned with regard to future risks*

ES – *no deberían plantearse ninguno de los problemas que **algunos de ustedes** han mencionado en cuanto a riesgos futuros*

FR – *aucun des problèmes que **certain d’entre vous** ont évoqués quant aux risques futurs ne devrait se poser*

PT – *não deve colocar-se nenhum dos problemas que **os senhores deputados** aqui referiram em termos de riscos futuros*

Action: P-block-alignment

EN – *Mr president, ladies and gentlemen*

ES – *Señor Presidente, señoras y señores*

FR – *Monsieur le Président, mesdames et messieurs*

PT – *Senhor Presidente, senhoras e senhores deputados*

Action: S-block-alignment of *Mr president* with its equivalents in the Romance languages

Action: S-block-alignment of *ladies and gentlemen* with its equivalents in the Romance languages, except for the non alignment of the Portuguese noun *deputados*, which is not in the other languages

EN – *Mr. Solbes*

ES – *Señor Solbes*
FR – *Monsieur Solbes*
PT – *Senhor Comissário Solbes Mira*

Action: S-block-alignment of the title and the surname *Mr. Solbes* in English, *Señor Solbes* in Spanish, and *Monsieur Solbes* in French

Action: P-block-alignment of the Portuguese title *Senhor Comissário* with the single title in English, Spanish and French

Action: P-block-alignment of the Portuguese double surname *Solbes Mira* with the single surname *Solbes* in English, Spanish and French

Action: S-segment-alignment of *Mr.* and *Solbes* with the equivalents in the Romance languages

6.6.2 Thanking

Thanking can be more or less formal. Some languages may use a more formal expression for thanking than others.

EN – *thank you very much*
ES – *muchas gracias*
FR – *merci beaucoup*
PT – *muito obrigado*

Action: S-block-alignment for the complete and more formal thanking expression (adverbial intensifier + thanking)

Action: internal S-block-alignment for *thank you* with *gracias*, *merci*, and *obrigado*