Guidelines for Arabic-English Word Alignment

Created by: Safa Ismael, email: safa@ldc.upenn.edu
Xuansong Li, email: xuangong@ldc.upenn.edu
With Contributions from:
Dalal Zakhary, email: zakhary@ldc.upenn.edu
Stephanie Strassel, email: strassel@ldc.upenn.edu
Stephen Grimes, email: sgrimes@ldc.upenn.edu

Table of Contents

1. Introduction

The word alignment task consists of finding correspondences between words, phrases or groups of words in a set of parallel texts. The resulting data can be used as gold standard training and testing data for machine translation. With references to Blinker and ARCADE project guidelines, these guidelines are especially designed to suit the task of Arabic-English word alignment. A visualized tool is developed by LDC to facilitate the task.

1.1 Overview of the Guidelines

The data used for word alignment is first presented in Section 2. In the Section 3, the tasks are specified, and the conventions adopted in the guidelines are explained in greater detail. In Section 4, the general strategies of annotation are addressed to deal with universal language features in word alignment. More detailed specifications and rules are illustrated with examples in Section 5. Section 6 handles unaligned words. Finally, Section 7 is devoted to approaches toward distinctive features of Arabic language.

2 Data

Data types include newswire, broadcast news, broadcast conversations (talk shows, call-in shows), newsgroup and weblogs.

Arabic words are tokenized automatically and then reviewed manually by annotators prior to alignment. Arabic tokenization follows the same guidelines used in Penn English Treebank: split words by white spaces, separate punctuations from the preceding/following words, apostrophes S ('s) are treated as separate tokens. Penn English Treebank treats most hyphens as separate tokens but some as part of words.

We tokenize the Arabic text by following the conventions of LDC Arabic Treebank tokenization: each Arabic word is treated as a separate token; English words in Arabic texts are tokenized according to Penn English Treebank standards; punctuation is separated; clitics are separated from the stem words. In Arabic word alignment, all hyphens are treated as separate tokens except when used in transcription to indicate a partial word.

3 Tasks and Conventions

3.1 Tasks

a) Link words or phrases in the source language (i.e. Arabic) to those in the target language (i.e. English).
b) Make judgments on translated and/or not-translated elements in the source and target languages.

c) Attach unmatched words to their related parts according to attachment rules.
d) Reject the alignment using the "Reject Segment" button for blank sentences, unmatched sentences, half translated sentences and pure English sentences on both sides.
e) Add a comment if any potential problems are spotted (optional).

3.2 Conventions

To better understand the word alignment annotation task and to facilitate the compilation of the guidelines, the following terms and symbols are employed.

a) Translated, not-translated, correct, incorrect are labels that appear in the tool for links and markups.
b) Brackets <> are used to indicate equivalence between the source language and the target language. E.g. president <> رئيس
c) Parenthesis ( ) are used to indicate emptiness or omission, E.g. `()<> the` This indicates "the" appears in the translation but no counterpart in the source language.
d) Colors, brackets, asterisks and plus signs in the guidelines indicate corresponding links and markups: brackets <> represent alignment, asterisks * represent attachment of such words, plus signs represent the split of tokens, blue highlights represent correct alignment links, yellow highlights represent "translated and correct" words, gray highlights represent "translated and incorrect" words, and pink highlights represent incorrect alignment links.

4 Concepts and General Approaches

For efficiency and accuracy, it is important to emphasize some general alignment strategies.
- The annotator should glance over both the source sentence and target sentence in the two windows on the right side of the tool interface to familiarize themselves with the sentence and ensure the translation is valid.
- Beginning with the source sentence, the annotator should first link all the content words before moving on to function word links.
- With all the equivalent links mapped, the left-over words or phrases can be either aligned to other parts or marked as *not-translated*.
- Most of the links are "translated correct" links; "not-translated correct" links are used for purely functional or grammatical word insertions due to language idiosyncratic features or for contextual insertions for the sake of effective communication and discourse coherence.
- All words in both source and target languages should either be linked or marked. No word should be left unattended. An annotator can reject a sentence if s/he thinks that sentence is not suitable for alignment.

4.1 Translated versus Not-translated

4.1.1 Translated

When translating from one language to another, it is common to see multiple translation versions for a particular word or a sentence in the source language. The different versions may all convey the same meaning, i.e., they are semantically equivalent but the difference may occur as a result of a word choice or style. Thus, each of these versions is a correct translation of the source words or phrases. Therefore, all these versions are considered to be "translated" and "correct". For mapping, an obvious lexical item or items can be found in the source and translation.

One source word, for example, can be translated in different ways (synonyms).

Example:
رفض عرض المدير <> (version 1) He refused the manager's offer.
(version 2) He declined the manager's offer.
(version 3) He said no to the manager's offer.

A pair is also treated as "translated" and "correct" even if there is a change in part-of-speech or sentence structure.

An example to show change of part of speech:
ليكن في قلب +ه الشفقة
Let his heart contain mercy

An example to show the surface structure change:
و+ اكد ابو عبيدة في تصريحات لــ « الشرق الاوسط » ، أن إسرائيل تتعمد تهويل و+ تضخيم قوة المقاومة ل+ إعطاء انطباع مضلل حول موازين القوى .
In statements to " Al-Sharq Al-Awsat , " Abu Obaida asserted that Israel intentionally exaggerates and overstates the resistance force in order to give a false impression about the balance of powers .

There are two types of links in terms of "translated" type:

1) Translated and incorrect

If a word or phrase is translated in a wrong way -- either semantically or grammatically -- they are treated as translated but incorrect. Typos or grammatical errors in the target language can be treated as "translated" and "incorrect".

An example to show semantically incorrectly translated:
حينما تشهد القدس على احداث العصر
When Jerusalem Witnesses the Events of the Afternoon

An example to show a non-grammatical translation:

وﻳﺟﺏ علـي كل الـشرفـاء والمـخلصين في الـعـالم مـؤازرتهم ولـو بـكلـمة حق

And all the honorable and loyal people in the world `support` them, even with a word of right

An example to show a typo in translation:

ونـقـل مـوقع مـيد بـاسين نـيـوز لايـن عن تـقـريـر أصدرتـه وزارة الخـارجيـة الأمـريـكيـة

`Mid Vasin` News Line website quoted a report published by the American State Department

2) Translated and correct

Translated and correct links are the default and are the most frequent type. The meaning is conveyed properly and they are grammatically correct. The links where direct equivalences of both form and content could be spotted are the easiest ones to detect.

An example to show obvious and direct equivalence between source and translation:

`و+` ذلك يسري علي الـسائقـين في لـوس أنجـلـوس و+ كـالـيفـورنـيا .

This applies to the drivers in Los Angeles and California .

A typo in the source side can still make a correct link to the translation when the translation is right.

An example to show a typo in the source:

و+ قـضت المحكة ب+ مـصـادرة أمـوال +هـما .

`The court` decided to confiscate their money .

4.1.2 Not-translated

A word is marked as "not-translated" when it is both semantically and lexically missing from the English translation or when the lexical representation of the word is missing from the translation but the semantic meaning is neither lost nor added. In such cases, extraneous words could be found. Thus two cases are recognized for "not-translated" markups:

1) not-translated and incorrect:

This kind of markup is proper when both the word form and the information are lost.

An example to show both word form and the information are lost in translation:

خيّم المئات من تلامذة الـثانـوية و+ معـلمين +ها و+ ادارييـن +ها ، و+
حتى الـعـامـلـين في+ ها ، طوال نهار امس
```
Hundreds of the high school students, teachers, and even staff,
"camped" all day yesterday.
```

In translation, we sometimes find the meaning of a word translated partially. In other words, we see partial concept entailment between the source word and translated word. We treat these words as "not-translated" and "incorrect" because partial meaning is lost anyway.

Examples to show word partial entailment in translation:

و+ على مقـربـة من +ه في قـاعة المسرح تـلامـذة يقـدمـون اولى الـبروفـات ل+
عمل مسرحي س+ يقـدم ب+ أكـثر من لـغة
```
Close by in the theater, students presented the first rehearsal of a
play that will be presented in more than one language.
```

و+ يَحظى الـبرغوثي ب+* شعبـيـة بـين الـفـلسطينـيـين المسلمين
```
Barghouti enjoys popularity among the Palestinians .
```

Topic related extra words, which are usually content words, are treated as "not-translated" and "incorrect" because the topic clue may be found within or beyond a sentence, and such inference may simply rely on reader's understanding of the discourse focus.

An example to show discourse-level content word insertions:

و+ كذّب هنية المزاعم الإسرائيـلية
```
Ismael Haniya refuted the Israeli allegations
```

2) not-translated and correct:

"Not-translated correct" refers to extra words inserted during translation. These words are either grammatically or contextually needed to make the translation fluent. For instance, the words "Holy قـد، وجـاء، في نهايـة المطاف،" in the following sentences are all extra words needed for Arabic and English to make the sentences grammatical and fluent.

و+ قـد فـر سيدي ولد سيدنـا من قصر الـعدل في نـواكشوط يـوم الاربـعـاء
في الـثـانـي مـن نـيـسان / ابـريـل
```
Sidi Wild Sidna fled from the Palace of Justice in Nouakchott on
Wednesday, April 2nd.
```

و+ جاء في بيان نشرته وكالة الانباء السعودية أقر المجلس عددا من
الخطوات على المدى القصير و+ البعيد في اطار استراتيجية متكاملة
```
In a statement published by the Saudi News Agency , the Council
approved a number of steps on the short and long terms as part of a
comprehensive strategy
```

و+ س+ يؤدي ذلك ، في نهاية المطاف ، إلى كارثة اقتصادية
```
And this will end up being an economic catastrophe .
```

ولد و+ ترعرع في مكة
```
He was born and raised in the Holy Mecca
```

Extra words can be inserted at the phrase level, the sentence level or a discourse level. For phrase (or local) level insertions, we use an attachment approach to align them to their related words (see Section 5). For other insertion cases not mentioned in section 5, we mark all of them as "not-translated correct" (see Section 6: unmatched/unattached words).

4.2   Minimum-Match

4.2.1   Minimum-Match in Literal Translations

In the guidelines, the principle of a word-for-word alignment is strictly followed, that is, the smallest number of words is preferred.

```
The following examples to show minimum match:
```

و+ ما هي الحاجات الأساسية في غزة الآن ؟
```
And what are the essential needs in Gaza now ?
```

و+ يعود الأمل إلى الناس
```
and hope can return to the people .
```

4.2.2   Maximum-Match in Idioms and Non-Literal Translations

There are cases where a minimum-match cannot be achieved. In such cases, to ensure a one-way semantic equivalence, annotators will need to choose as many words as necessary for one link. That is, select as many characters as you can to make a semantically independent unit. Such cases include formulaic or frozen expressions, idioms (or near idioms), proverbs and hyphenated words. Other cases include very tricky structures such as verb phrases (verb particles) and words with affixes.

The following examples to show the alignment of idioms, fixed expressions, and proverbs:

قـبـل أن يـنـبـت الـشعـر في ذقـن ه+
Before he reaches puberty .

أسعد الله ب+ الخير مـساء كم+
Good evening .

الـطيـور عـلـى أشكـال ها+ تـقـع
Birds of a feather flock together .

إنَّ ه+ ل+ مـن دواعـي سرور ي+
It is my pleasure .

For proverbs, however, if a word-for-word translation is found, sub-part links would be preferred especially for those directly borrowed from English.

يضرب عصفـوريـن ب+ حجر واحد
Kill two birds with one stone .

الـكلاب الـنـابـحة قـلـمـا تـعضّ
Barking dogs seldom bite .

We follow the same approach for proper nouns, which are described in Section 5.5.

Non-literal translations -- other than idioms or proverbs -- are also very common. In such cases, drilling down to individual constituents would result in semantically non-equivalent alignments. To avoid such non-equivalent alignments, we would prefer to treat them as whole units.

دمت بـعـافـيـة
May God keep you healthy .

Hyphenated words in English are treated as one unit if the subparts of the hyphenated words are inseparable. In tokenization, these inseparable words are treated as one word. For those which are not tokenized as one word, we will not break down into subpart alignments.

An example to show hyphenated words are aligned as a whole unit:

آخر مـا تـوصل إلى ه+ الـعلـم
State-of-the-art

4.3 Attachment Approach

Extraneous words on the phrasal level are usually attached to their dependency constituents when they are grammatically or contextually required (see Section 5 on attachment rules).

We align extra words to their generating heads in complex situations. Such words can be conjunctions or prepositions.

Examples: the highlighted words should be aligned to their heads**:**

الحكومة الإسرائيلية بدأت منذ قليل <mark>اجتماعا</mark> ينتظر أن يسفر عن إعلان <mark>وقف إطلاق النار</mark> .

```
The Israeli government has recently started a meeting that* is expected
to result in the announcement of* a ceasefire .
```

<mark>و+* هو</mark> ما يشكك <mark>في* +ه</mark> عدد من المحللين .

```
This is* what a number of analysts doubt .
```

5   Alignment and Attachment Rules

5.1   Anaphora (pronouns)

Anaphora are the co-reference of one expression with its antecedent. The links and markups can be done as follows.

Case 1: Pronoun <> Pronoun

أخبر +ني جدّ <mark>+ي</mark> حكاية
My grandfather told me a story

Case 2: Pronoun <> Verb

هذا الشيخ <mark>هو</mark> جدّ +ي
This old man <mark>is</mark> my grandfather.

In Arabic, a subject is sometimes dropped after its first mention whereas it is kept in English. Although omitted, the Arabic subject is understood from the verb. We attach the pronoun or subject to its referents and consider them as "translated" and "correct".

The following examples to show the omission of subject:

Case 3: Verb <> Pronoun+Verb

قاد الصحاف حملة إعلامية مغلظة حين <mark>كذّب</mark> وجود الإحتلال

Al- Sahhaf lead faulty media campaign when `he*` `refuted` the presence of the occupation.

قـاتـل في مـعـارك جّمة
`He*` `fought` in several battles

Case 4: Pronoun <> Referent

حسّان أخاً و+ صديقـا . هو لم يـتأخر يـومـاً عن مـساعدة +ي
Hassan is a friend . `Hassan` `was*` never late to help me

Case 5: Referent <> Pronoun

عـند وصـول هذا الأخير
Upon `his` arrival .

We adopt the same attachment approach for reflexive pronouns (myself, yourself, herself, himself, itself, ourselves, yourselves) and indefinite pronouns (one, some, someone).

5.2 Demonstrative Words

Links involving demonstratives (this, that, these, those, here, there) can be illustrated by the following cases.

Case 1: Demonstratives <> Demonstratives

هنـالك حشد مـن المـتظاهريـن ب+ الـقـرب مـن هنا
There is* a crowd of demonstrators nearby

هؤلاء هم الـنخبـة
Those are the elite .

Case 2: Demonstratives <> Referent (or vice-versa)

إقـترح فـريـق تـرسيم الحدود ل+ لـدولـة . ف+* هم مـن المـتخصصين
A team suggested rectifying the border of the country. The team is* specialized .

As illustrated in Case 2 above, demonstratives with no translation equivalent are attached to their head word(s).

5.3 Copula BE

Link verb "to be" (am, is, are, be, been, being) qualifies or provides information about the properties of its argument. Many languages do not use the verb "to be" to separate the subject from its object, such as in Arabic, then in link mapping, the extra link verb can be treated as "not-translated" and "correct".

An example to show "be" is aligned:

الـذي نـفـذ الـعمليـة الإستشهـاديـة <mark>هو</mark> عمر سلمـان
The one who carried out the martyrdom operation <mark>is</mark> Omar Salman

Examples to show "be" are omitted:

نحن أقـويـاء
We <mark>are</mark> strong .

هذه المغامـرة تـكلف +نـا الـكثير
This adventure <mark>is</mark> costing us a lot .

5.4 Proper Noun

A proper noun is used to name a specific item. It is usually a one-of-a-kind item and it always begins with a capital letter.

We follow the basic minimum-match approach in aligning proper nouns, especially when they are compositional (usually one or more specific name plus one or more common nouns).

<mark>عبـد الـرحمن</mark> الـكـواكبي
<mark>Abdulrahman</mark> Al-Kawakibi

<mark>أم كـلـثوم</mark>
Um Kulthoum

سيدي أحمد <mark>بن بيلا</mark>
Sidi Ahmed <mark>Ben Bella</mark>

For names of people, we align first name and last name separately.

<mark>مـيـسون</mark> الـدمـلـوجي
<mark>Maysoon</mark> Al-Damalooji

محمـود <mark>عبـاس</mark>
Mahmoud <mark>Abbas</mark>

For names of countries and such entities in their abbreviated forms, we treat them as whole units.

<mark>الـبـيت الأبـيض</mark>
<mark>The While House</mark>

When the name is not abbreviated and has good one-to-one match, we'll break down to sub-piece alignment.

الـولايـات المتحدة الأمـريـكـية
The United States of* America

Acronyms of translation of proper nouns are treated as whole units.

الـنـاتـو
North Atlantic Treaty Organization

أنـروا
United Nations Relief and Works Agency

However, if there is no perfect match for the case of abbreviations for proper nouns, then we align them as a whole unit.

يـو إن إتـش سـي آر
UNHCR

5.5 Determiners (articles in English)

Determiners -- particularly articles in these guidelines -- are used before nouns to indicate whether one is referring to a specific object or a particular type. The definite and indefinite articles (a/an/the) are all determiners. Such determiners are attached to their head nouns. In Arabic, however, there are no indefinite articles but there is only one definite article, "الـ". It always comes attached to the noun so we align it together with the noun.

Sentences involving articles can have the following cases:

Case 1: Determiner <> Determiner

الـقطـار الـسريـع
The express train

Case 2: (none) <> Determiner (or vice-versa)

شربـت قـنـيـنة كبـيرة مـن الـعصير
I drank a large bottle of juice .

الـطفـل الـصغير
The little child

It is common to have a determiner in one sentence that is omitted in the translation. We align the extra determiners to their head noun even if the head noun is not adjacent to the determiner.

Case 3: Determiner <> different translations (or vice-versa)

وأتــى <mark>الهجوم</mark> في وقت كــان الـقـلـق في +ه يــتـزايـد
<mark>This attach</mark> came in time where tension was at its utmost .

It happens occasionally that a determiner in one sentence segment has a non-determiner translation equivalent.   In this case, the alignment is created according to the meaning expressed.

5.6   Auxiliary Verbs

An auxiliary or helping verb is a verb giving further semantic information about the main or full verb following it. An auxiliary verb can have many forms. Auxiliary verbs are used to serve such functions as passive, progressive, perfect, modal, or dummy. Whenever the auxiliary verbs appear in both sentences, simply align them. When missing in one sentence, the extra auxiliary verbs will align to the main verb.

An example to show auxiliary used together with negative words:

<mark>لم يــكن</mark> يــدري أن الـوقت قـد حـان ل+ الـعودة
He <mark>was</mark> <mark>not</mark> aware that it was time to go back

An example to show how extra auxiliary verbs are aligned:

تــكـاد الأزمـة الـلـبـنـانـية <mark>أن</mark> تـتجمد
The Lebanese crisis reached the freezing point .

The last example is just an exception because "n't" is connected to the auxiliary verb. Normally, "n't" should be tokenized as a separate token.

The word "do" may be used for "emphasis"; in this case it should be not translated correct.

5.7   To- Infinitive

In Arabic-English alignment, "to" infinitive is aligned withthe verb following it. Occasionally, some words might be inserted between infinitival to and the verb, a case which requires extra attention to make a correct alignment.

An example to show words are inserted between an infinitive "to" and the verb:

هذا قـرار ربـا علــى +ه <mark>إتخاذ</mark> +ه
This is a decision he has <mark>to</mark> possibly <mark>make</mark> .

Occasionally, "to" infinitive may find its translation equivalent, where a direct link would be possible.

An example to show "to" is aligned:

كــان علــى المجرم <mark>أن</mark> يــعترف
The criminal had <mark>to</mark> confess

5.8   Expletives

Here "expletive" mainly refers to syntactic expletives: words that perform a syntactic role but contribute nothing to meaning, such as "it", "there" "here". For cases where an equivalent counterpart can be found, we link them. If without equivalence, the extra "it", "there" or "here" can be marked as "not-translated" and "correct".

Those examples are of unaligned expletives:

ذات مــرة
<mark>There</mark> comes a time

<mark>قــد</mark> حــان وقت الــعشاء
It is time for dinner

<mark>إن</mark> الجهل ظلام
Ignorance is darkness

5.9   Conjunction

5.9.1   Conjunctions with "and"

Four cases of conjunctions are involved under this category:
  a) و <> and
  b) , <> , and   (or و ,<>,)
  c) و <> ( ) (or () <> and)
  d) () <> و

For case (a), simply link both sides.

<mark>+و</mark> رغم الــوساطات المتتابــعة خرجت المعــارضة الى الــشارع
<mark>And</mark> despite the successive mediations, the opposition went into the street .

For case (b), link comma to comma, the extra "and" is treated as "not-translated correct".

الـوضع الإقـتصادي، والـذي بـلغ حجم ديـن +ه الخـارجي 40 مـليـار دولارا
يحـتاج الى تـوافـق وطني حول الإصلاحـات الإقـتصـاديـة
```
The economic situation, and that has reached 40 billion dollars, needs
a national accord on economic reforms .
```

For case (c), link comma to "and"

فـ+ زار المكان، حيّا أهل +ها، ثم عـاد إلى الـسيـارة
```
He visited the place and greeted its people, to later return to the
car .
```

For case (d), the extra "and و" is marked as "not-translated" and "correct".

و+ بـ+ مجرد أن إنـتهى الـوزيـر مـن حديث +ه
```
As soon as the minister finished his statement
```

5.9.2 Conjunctions without "and"

Conjunctions other than "and" are illustrated in the following example:

Example of "neither… nor", "either…or":

تـركة مـايـكل جـاكسون لم تـكن بـيضاء او سوداء
```
Michael Jackson's legacy is neither black nor white
```

5.10 Prepositions

For prepositions, if there exist equivalent constituents, simply link the correspondent parts.

For cases of two or more propositions in English but only one in Arabic, glue the two prepositions and align them to the preposition counterpart in the other side. For cases of one preposition on one side while no counterpart on the other, the extra proposition will be glued to the object of this preposition even if the object is far away.

Example to glue extra preposition:
باب الصف مكسور
The door of* the class is broken

For cases where the extra preposition appears after a verb, it is glued to its object despite an obvious co-occurrence of the verb and the preposition. This same rule holds for a preposition following a noun.

Example to glue extra preposition after a verb:

إرتفع عدد الضحايا <mark>ل+* يصل</mark> إلى سبعة آلاف
The number of victims <mark>has reached</mark> 7000

When the object of an extra preposition is a present participle, align the preposition to the participle.

An example to glue extra preposition to the participle:

لقد قبض <mark>على* القاتل</mark>
<mark>The killer</mark> was arrested

5.11 Phrasal Verbs

Phrasal verbs are verbal complexes comprised of a verb and a particle. The particle may be a preposition or an adverb. As the particle-verb combination crucially alters the semantics of the verb, we treat them as a unit for this task (i.e, align the particle with the verb).

Example of verb particle alignment:

الوزير <mark>أشار إلى</mark> أهمية التعليم
The minister <mark>pointed out</mark> to the importance of education .

5.12 Possessives

English possessives can take three forms: "**of**", "**s**" and " **'** ". The Arabic nominal complementation structure is considered the equivalent of the English possessive construction: "of", "'s" and " ' ". If possessives appear on both sides, simply link them.

Examples to show the alignment of "'s":

ضاعت حقيبة <mark>الطفل</mark>
The <mark>child's</mark> bag has been lost

كانوا يحتفلون ب+ العيد الأول <mark>ل+</mark> زواج +هما
They were celebrating their wedding<mark>'s</mark> first anniversary .

An example to show the alignment of "of":

إنفجر إطار <mark>السيارة</mark>
The tire <mark>of* the car</mark> exploded

English plural possessive consisting of only an apostrophe are treated the same as an apostrophe plus "s" sequence because they are semantically equivalent.

An example to show the alignment of " ' ":

```
مفاتيح الابواب
Doors' keys
```

For extra 's where no equivalent can be found in the source, they are marked as "not-translated correct".

This example is to show extra "possessive" alignment:

```
إنتقدت " هآرتز " أمس في إفتتاحية +ها القرار الذي إتخذت +ه
اللجنة المركزية في ليكود
The opening page of yesterday's Haaretz criticized the decision taken
by the Likud central committee .
```

5.13 Passive Sentences

If passive voice appears in both source and target languages, it might be easy to link each piece correspondingly. Whereas for the case of passive voice in the source and the active voice in the translation or vice versa, the word order would be very different and it is not easy to tear apart the corresponding parts. Great care is needed in correctly mapping the links.

Examples to show how to align different cases of passive voice sentences:

```
إن مسودة الدستور كتبت ب+ تدخل أمريكي بريطاني مباشر
The draft of the constitution was written directly by American and
British interference .
```

```
طُلب من +ه المثول أمام المحكمة
He was asked to appear before the court .
```

5.14 Negation

Negation in Arabic is expressed by different forms using different particles including: ما، لا، لم، لن، ليس. There are often direct translations of the Arabic negation into English.

```
لا يوجد عند +ي شئ أضيف +ه
I do not have anything to add .
I don't have anything to add .
```

```
لم أقل ذلك
I did not say that .
I didn't say that .
```

إذا أنت <mark>لم</mark> تشرب مراراً و+ <mark>لم</mark> تتجرع مرارة +ه ضمئت
If you <mark>have not</mark> had impure water , or taste its bitterness , you get
thirsty .

لا <mark>يعقل</mark> أن تقول ذلك
You <mark>can not</mark> say that .

5.15 Relative Nouns

Relative nouns in Arabic relate a subordinate word or clause to the rest of the
sentence:(الذي، التي، اللذان، اللتان، اللذين، الذين، اللتين، الألى، اللاتي، اللائي). In English there
are five relative pronouns: that, which, who, whom, and whose. For Arabic
relative pronouns, we align them to their English counterparts.

a) Relative noun <> Relative pronoun

الإعترافات <mark>التي</mark> أدلى ب+ ها سابقاً و+ أدت إلى إتهام +ه ....
The confessions he had previously given <mark>which</mark> led to him being
accused...

كل <mark>ما</mark> تتهم +نا ب+ ه إسرائيل سواء كان حقاً أم لم يكن
Everything of <mark>which</mark> Israel accuses us, whether it is true or
otherwise..

b) Pronoun <> Relative pronoun

في البلدة ضجة يتنافس على تسعير +<mark>ها</mark> ، من فوق ، طرفان لا يحظيان
ب+ الثقة الوزارة من السواد الأعظم من الناس
There is a clamor in the country <mark>which</mark> two parties, who lack the
balancing confidence of the vast majority of the people, are trying to
kindle from above .

c) Relative noun <> (none)   and vice-versa

Sometimes those pronouns appear in the Arabic source and not in the English
translation and vice versa. There are three cases:

Case 1: The relative noun or pronoun refers only to one word. We glue them to
their antecedent when it is only one word, and it would be unmatched and labled
with the "glue" tag.

لذلك نحن نفهم <mark>الجهود التي*</mark> تبذل +ها الوزارة
We thus understand <mark>the efforts</mark> being made by the ministry .

الصين تعمل على تطوير <mark>صواريخ</mark> فضاء تقرّب +ها من حلم +ها
China is working to develop space <mark>rockets that*</mark> will bring it closer to
its dream .

Case 2: The relative pronoun refers to more than one word. We treat the relative pronouns as "not translated and correct", applying the minimum-match strategy.

```
و+ كان الرئيس اللبناني السابق اميل لحود و+ الرئيس السوري
بشار الأسد اللذان أمرا ب+ إغتيال الحريري
The former Lebanese President Emil Lahud and Syrian President Bashar
Assad had offered the assassination of Hariri .
```

5.16 Subordinate Clauses

When the clauses are placed in a hierarchical structure, the more important ones are main clauses and the less important are called subordinate clauses since they cannot stand alone as sentences because they do not provide a complete thought.

A subordinate clause in English is introduced by a subordinate conjunction or a relative pronoun and contains both a subject and a verb. Subordinate conjunctions include *after, although, as, because, before, even if, even though, if, in order that, once, provided that, rather than, since, so that, than, that, though, unless, until, when, whenever, where, whereas, wherever, whether, while, and why.*

Depending on the environment, subordinate clauses can function as subjects, objects, complements and adverbials.

a) clause as subject
   The extra clause markers are treated as "not-translated correct":

```
رغم أن +ني لم احب محاضرة +ه ، ف+ كان على +ي الإستماع
Even though I was not interested in his lecture, I had to listen
to him .
```

```
ما لم تغير من اسلوب +ها مع الناس ، ف+ إن +ها لن تكسب
ثقة +هم
Unless she changes her attitude with people , she will not gain
their trust .
```

b) clause as object:

```
لم يجتاز عليّ إمتحان التاريخ لإن +ه قضى الليلة مع صاحب +ه في
المشفى
Ali did not pass his history exam because he spent the night with
his friend at the hospital .
```

5.17 Punctuation

All equivalent punctuations can be aligned, including comma-to-comma, period-to-period, and so on. Extra punctuations can be marked as "not-translated" and "correct". Special cases in Arabic are:

Case A: : <> ,

:وقـال <> He said,

Case B: , <> . (comma on one sentence, period on the other). For this case, link comma to period:

و+ إستـمـر فـي كـلام +ه ، حيث أضـاف ...
And he continued talking . He then added ….

Case C: "and" <> "," (or vice versa)

see Section 5.9.1 Conjunctions.

Case D: (none) <> various punctuations (or vice versa)

قـبـل إنـعـقـاد الإجتـمـاع كـان ل+ نـا لـقـاء مـع الـوزيـر
Before the meeting was held , we met with the minister .

For cases where there is a punctuation mark in one sentence while it disappears in the other sentence, the extra punctuation mark is treated as "not-translated" and "correct".

6   Special Features in Arabic

6.1   Interrogative Sentence

6.1.1   Usage of "Hamza" and "Hal"

The Arabic interrogative words هل and أ have variable equivalents in English language. The question is most of the time asked in English by varying the word-order of the subject and predicate, or by using function words:

هل مـعـنى هذا أن لـغة واشنطن قـد أفـلـحت ؟
Does this mean that Washington's language has succeeded ?

هل أولاد +كم مـدلـلـون ؟
Are you children spoiled ?

هل <mark>يمكن</mark>+ك أن تجلب الكتاب ؟
<mark>Can</mark> you bring the book ?

<mark>أ</mark>+ راغب أنت ؟
<mark>Are</mark> you willing ?

6.1.2 Interrogative Negation

In cases where you have both interrogation and negation at the same time, we link those which have direct translation in English.

<mark>أ</mark>+ لم ترى +ه ؟
<mark>Did</mark> you not see him ?

<mark>أ</mark>+ ليس من المعقول أن نطالب ب+ الحرية ؟
<mark>Would</mark> it* not be logical to demand freedom ?

6.2 VSO versus SVO Structure (subject-drop)

The constructs of "verb before subject" (VSO) and "subject before verb" (SVO) are common structures in Arabic even though classical Arabic tends to prefer VSO. SVO is more common in spoken Arabic, or rather in colloquial Arabic and in less formal MSA (Modern Standard Arabic), and case endings are dropped in pausal forms.

Examples of SVO and VSO alignment:

<mark>الطفل</mark> أكل التفاحة
أكل <mark>الطفل</mark> التفاحة
<mark>The child</mark> ate the apple .

Subject pronouns are normally omitted in Arabic because the subject is marked on the verb.

Example of alignment in the case of a null subject:

ذهبت إلى المنزل للغداء
I went home for lunch .

Note, however, that the subject may be used for emphasis or when using a participle as a verb where the participles are not marked for person. In this case the subject pronouns are aligned to each other.

<mark>أنا</mark> قلت ذلك مرارا
<mark>I</mark> said that repeatedly .