

# The challenge of adapting PISA materials into non Indo-European languages

Some evidence from a brief exploration of language issues in Chinese and Arabic

Doc: TAG(0801)

August 2007



Core A Consortium:

Australian Council *for* Educational Research (ACER)

Unité d'analyse des systèmes et des pratiques d'enseignement (aSPe)

cApStAn Linguistic Quality Control

Deutsches Institut für Internationale Pädagogische Forschung (DIPF)

National Institute for Educational Policy Research (NIER, Japan)

Westat

## Table of contents

INTRODUCTION.....	3
POTENTIAL LINGUISTIC PROBLEMS IN ARABIC.....	6
Standard Arabic <i>vs</i> spoken Arabic.....	6
Sentence length.....	6
Quantifiers.....	6
Numerals.....	7
Direct speech.....	7
Use of wh- words.....	7
Reference chains.....	7
Impersonal and generic expressions.....	7
Definite and un-definite articles.....	8
Past tense.....	8
English use of past tense in scientific texts.....	8
Plural.....	9
Place of adverbs.....	9
Grammatical structure in multiple-choice items.....	9
Vocabulary.....	10
Idiomatic expressions.....	11
Foreign words.....	11
Tables and graphics.....	12
Abbreviations.....	12
Use of capital letters.....	12
The “Explain your answer” issue.....	12
English shortcuts.....	12
Cultural sensitivity issue.....	13
POTENTIAL LINGUISTIC PROBLEMS IN CHINESE.....	14
Word order.....	14
Plural.....	15
English plurals used to convey a generic meaning.....	15
Plural or singular? The “Which of the following?” issue.....	15
Tense.....	16
Verbal modalities.....	17
Use of wh- words.....	17
Comparative expressions.....	17
Numerals.....	18
Pronouns.....	18
Vocabulary.....	19
Literal matches involving derived forms of a same word.....	19
Idiomatic expressions.....	19
Answer categories.....	20
Use of capital letters.....	20
Foreign names.....	21
Abbreviations.....	21
The “Support your answer” problem and other issues with instructions to students.....	21
English shortcuts.....	22

## INTRODUCTION

To what extent was equivalence achieved for the Reading, Mathematics and Science test materials used in the PISA 2000, 2003 and 2006 studies, across all national translations into the various languages of instruction of the participating countries?

To explore this issue, indicators of international equivalence of item difficulties were created for each of the three domains, using the communalities obtained from PCA analyses, where the Rasch estimates of item difficulty at the national level (centred deltas) were used as observations and the national versions as variables<sup>1</sup>.

In Mathematics, the average communality was no less than 91% of the cross-country variance of item difficulties, indicating a very substantial level of equivalence. The value was also high (around 80%) for Reading and for Science.

However, in all three analyses, the indicators of equivalence appeared to have significantly lower values for a number of national versions, most of which were developed by Middle East and Asian countries using non Indo-European languages (see Table 1).

In particular, the communalities were lower than desirable for all four Arabic versions used in Tunisia, Jordan, Qatar and Israel (and, to some extent, for the Hebrew version used in Israel and the Turkish version used in Turkey). The values were particularly low for the national versions in Azeri, Kyrgyz and Uzbek used in Azerbaijan and Kyrgyzstan (but not for the Russian versions also used in these two countries). As regards Asia, the indicator was low for the Japanese, Korean, Thai and Bahasa Indonesia versions, and for the three Chinese versions used in Taiwan, Hong Kong and Macao. Interestingly, no such problem was observed for the versions in non Indo-European languages used in Western and Central Europe (Finnish, Estonian, and Hungarian), except perhaps for the Basque version used in Spain.

This issue did not seem to be specific to PISA. Parallel analyses conducted on TIMSS and PIRLS item difficulties also indicated a trend towards lower levels of equivalence for versions in non Indo-European languages, particularly the versions in Asian languages, in African languages and in Arabic.

---

<sup>1</sup> See Grisay, A., de Jong J.H.A.L., Gebhardt, E., Berezner, A., Halleux-Monseur, B. (2007). Translation Equivalence across PISA countries, *Journal of Educational Measurement*, 8(3), 249-266, and Grisay, A and Monseur, C. (2007) Measuring the equivalence of item difficulty in the various versions of an international test, *Studies in Educational Evaluation*, 33 (1) 2007, 69-86.

**Table 1. Mean communality of item difficulties in various groups of national versions of PISA instruments**

**Indo-European versions**

**Overall**

	<b>Reading</b>	<b>Maths</b>	<b>Science</b>
Mean communality	<b>0,852</b>	<b>0,920</b>	<b>0,835</b>
Std	0,068	0,030	0,029
N versions	36	37	43

**Non Indo-European versions**

**Overall**

	<b>Reading</b>	<b>Maths</b>	<b>Science</b>
Mean communality	<b>0,708</b>	<b>0,816</b>	<b>0,722</b>
Std	0,109	0,072	0,081
N versions	9	12	17

**Asian languages**

	<b>Reading</b>	<b>Maths</b>	<b>Science</b>
JPN_JAP	0,664	0,777	0,648
KOR_KOR	0,690	0,797	0,701
HKG_CHI	0,685	0,806	0,758
MAC_CHI		0,835	0,754
TWN_CHI			0,632
IDN_IND	0,489	0,709	0,664
THA_THA	0,672	0,844	0,683

**Middle-East languages**

	<b>Reading</b>	<b>Maths</b>	<b>Science</b>
ISR_HEB	0,804		0,728
ISR_ARA			0,760
JOR_ARA			0,604
QAT_ARA			0,667
TUN_ARA		0,691	0,556
TUR_TUR		0,847	0,759
AZE_AZE			0,545
KGZ_KYR			0,565
KGZ_UZB			0,553

**Non Indo-European languages used in Western and Central Europe**

	<b>Reading</b>	<b>Maths</b>	<b>Science</b>
ESP_BAQ		0,830	0,761
EST_EST			0,820
FIN_FIN	0,823	0,924	0,824
HUN_HUN	0,851	0,934	0,809

This paper presents the results of a limited study conducted with the help of cApStAn to identify possible issues that might explain the larger linguistic differences observed in these

groups of countries, with a view of providing a wider range of recommendations about potential translation traps in the *PISA Translation/Adaptation Guidelines*.

Due to time and practical constraints, only two non Indo-European languages, Arabic and Chinese (both used in several countries participating in PISA) were selected for this study.

A set of three Reading units from the first bundle of PISA materials being prepared for the P2009 Field Trial was submitted by cApStAn to a team of professional Arabic and Chinese translators, most of whom had been involved in previous PISA translation and/or verification exercises and were familiar with the materials and the *Translation/Adaptation Guidelines* used in the study. They were asked to double-translate these units using the recommended PISA procedures. Then the materials were reconciled, respectively, by a third experienced Chinese translator and copy writer and by two Arabic experts, all familiar with translation issues in international studies.<sup>2</sup> Both the translators and the reconcilers were requested to pay special attention to any translation problems encountered during their work, and to document them extensively (in English) in an Excel spreadsheet specially prepared for the purpose of the study. Their remarks are summarised in the two sections below.

Despite the limited scope of the study, the resulting picture was extremely interesting. A surprisingly large number of very diverse translation challenges were identified in the three units used in the exercise, many of which can probably also be of interest for translators in other languages than Arabic or Chinese

---

<sup>2</sup> The Arabic translation team was composed of Rachid Akel from Lebanon (a Mashreq country) and Elmosthapha Benboucetta from Morocco (a Maghreb country) – both professional translators, who were involved in the development of a generic Arabic version for TIMSS 2007. The two Arabic experts involved in the reconciliation of the three units were Hamzeh M. Dodeen, from Jordan, chair of the Program of Outcome Assessment at the University of United Arab Emirates, and Lina Daouk from Lebanon, a psychometrician teaching at the Assessment department of the Cambridge University.

The Chinese team was composed of two professional translators from Chinese Taipei, Yang Nanqian and Meichou Chou. The reconciler was Tsuchin Wu, also a professional translator from Taipei, a graduate in Journalism and International Relations. Tsuchin Wu was involved in the verification of the Chinese versions of the PISA 2003, PISA 2006 and IEA/TEDS-M studies.

## POTENTIAL LINGUISTIC PROBLEMS IN ARABIC

### Standard Arabic vs spoken Arabic

One of the important problems that exist in translation into Arabic is the fact that spoken Arabic is very different from official or written Arabic. Quite often, translators are unable to reproduce the exact English meaning in everyday Arabic language and they translate it into a difficult, non-common Arabic language. In this small exercise, the language used in both the Arabic translations received did not always sound “natural”. For example, in the sentence “*The change in sleep pattern occurs almost 18 months earlier for girls than for boys*” (R404 Sleep) both translators used difficult expressions that made the passage unintelligible for Arabic students. The reconciler reworded the sentence into “*Sleep patterns begin to change almost 18 months earlier for girls than for boys*”.

The common word for “TV” is “Television” in Arabic, while the official translation is “Telefaz”, which is not commonly used. For the sake of simplicity and consistency, the reconcilers recommended using the former in R407 WldChp, for all occurrences of the abbreviation “TV”.

### Sentence length

Long sentences are more acceptable in Arabic than in English, so translators may be tempted to collapse two or more short sentences into a single long one. In R405 Ainu, one of the reconcilers considered as too short the 3 sentences in “*I was happy living with my family. As the years passed, people began to call me Grandfather. I no longer went into the mountains to hunt.*” She suggested grouping them into a single sentence, as she did for a number of other passages in the texts used for this translation exercise.

This can only be done if (i) the difficulty of the resulting sentence is not increased for the students, and (ii) there is no specific question in the unit that refers to one of the sentences. For instance, if the unit R405 Ainu contained a question such as “*Quote a sentence from the text to show that the Ainu elder was loved and respected by the people in his village*”, this question would become more difficult for Arabic students in case the 3 sentences mentioned in the example above were combined into a single one, because the relevant sentence (“*people began to call me Grandfather*”) would be included into a much longer context than in the source version, which could probably affect the functioning of the item.

### Quantifiers

Certain English quantifiers are sometimes considered as unnecessary or difficult to render into Arabic. Both translators omitted the word “any” in “*Are there any tell-tale biological signs that youth is over and adulthood has begun?*” (R404 Sleep). One of them dropped the words “very much” from “*The transition from youth to adulthood is very much an interplay between social, psychological and physiological factors*” (R404 Sleep), while the other translator and the reconcilers preferred to use “*in fact*” rather than “*very much*”.

In the question stem “*What is one difference between this sentence and the rest of the story?*” (R405 Ainu), the word “one” was omitted by one of the Arabic translators and mistranslated

as “*the only one difference*” by the second. The reconcilers corrected into “*one difference*”, but recognized that this sounded slightly awkward in Arabic.

## **Numerals**

Although one of the translators considered as more appropriate to write numbers as words when translating age information (e.g., in R404 Sleep she adapted “*20.9 years of age*” into “*twenty years and nine months*”), both reconcilers recommended to use Arabic numbers (1, 2, 3, ...) for all numerals, first because they are universally understood, and also because there are problems when including decimals in Indian numbers used in Arabic when translating numbers into words.

## **Direct speech**

When using direct speech in English, such as in “*‘Hey, you Ainu people! (...) Hear my side of the story’, the fox said sadly, with tears in his eyes*” (R405 Ainu), the reference to the person who speaks can appear either at the beginning, or at the end, or within the reported discourse (in this specific case, at the end). The Arabic usage would rather place “*The fox said sadly*” at the beginning of this long declamation. An alternative solution, chosen by one of the translators, was to add a period at the end of the quotation, and to reword the reference to the speaker into “*This is how the fox ended his speech on a sad note and with a tear in the eye*”. Both reconcilers preferred to place “*The fox said sadly*” at the beginning of the passage.

## **Use of wh- words**

The expression “*whether or not*” has no direct translation in Arabic. When translated literally, the sentence “*Whether or not one is regarded as an adult depends partly on cultural circumstances*” (R404 Sleep) would become something like “*True or false: one is regarded as an adult...*” A better solution is to place the negation next to the word “*adult*” so that the sentence reads “*Considering one as adult or not adult depends...*”

## **Reference chains**

When several pronouns are used in the same sentence, such as in “*He would have liked to look at him*” (R407 WldChp), a literal Arabic translation may be ambiguous, meaning either “*he wished he would look at his opponent*” or “*he wished that his opponent would look at him*”. In such cases, it is preferable to replace the second pronoun with a synonym.

## **Impersonal and generic expressions**

Translating into Arabic the English word “*one*” is not easy when it is used in impersonal expressions such as “*Whether or not one is regarded as an adult*” (R404 Sleep). The most common translation in this case is “*the one*”, meaning “*the person*”.

“*An Ainu*” is used in English with a generic meaning in “*I am an Ainu*” (R405 Ainu). This may be ambiguous in Arabic, so the translators preferred a more explicit expression “*I am from the people of Ainu*”.

In casual English, the pronoun “you” can be used as a generic or indefinite pronoun referring to an unspecified person (e.g. “somebody” or “a person” or “people”), but in a way that is less formal than the use of the generic pronoun “one” and is still meant to somehow involve the reader. Compare, for example the instruction “After using this product, one should always wash his hands” with “After using this product, you should always wash your hands”. The use of generic “you” appeared to raise a problem for Arabic translators, in R407 WldChp: “Just as you clasp each other when you know that the next day each of your limbs will be burning, and that for some weeks to come you will still be counting the blows”, mainly because “you” can be both singular or plural in English, while it has to be explicitly marked for number in Arabic. One of the translators used dual in “you clasp each other” (suggesting that this part of the sentence referred to both opponents in the story), but singular for “you will still be counting the blows” (suggesting that here the world champion is only speaking to himself). The reconciler used dual in the two sentences, suggesting that both referred to the two opponents mentioned in the text. Anyway, the generic meaning seemed to be lost in each of these alternative interpretations.

### Definite and un-definite articles

In English, definite articles can be used in generic expressions referring to a *category* of objects or a *species* of animals, such as in “The Ainu did not create the salmon. Of course, the fox did not create salmon, either” (R405 Ainu). In both sentences, “salmon” will be turned into plural in Arabic in order to refer to the species rather than to a specific salmon or fox.

### Past tense

Arabic has two main tenses: perfect tense, referring to *completed* actions and imperfect tense (or present) referring to actions that are *still continuing*.

In “And then he had become a world champion” (R407 WldChp) perfect tense will be used. However, in the sentence that follows immediately the previous example, “And he was a world champion and he was thinking of the five hundred million TV viewers watching him” the first verb “and he was a world champion” can’t be in the perfect tense in Arabic (which would suggest that “he was a world champion before that match”). It needs to be put in the imperfect/present tense in Arabic to render the English nuance of present-in-the-past (i.e. “And at that point in time he had become a world champion”. In the rest of the sentence (“he was thinking of...”), perfect tense should be used to indicate that the whole situation is situated in the past.

### English use of past tense in scientific texts

Past tense is commonly used in English when reporting the results of a study, e.g. “For each person, Professor Rönneberg estimated the average time in the middle of the sleep periods. He **found** that this **got** later and later as the children **got** older until they **reached** about 20 years of age. At that age, the tendency **was** suddenly reversed, as the people in the study **started** going to bed earlier and getting up earlier” (R404 Slep).

In a number of other languages, including Arabic, it may be more appropriate to switch to present tense when reporting *generalisations* from the study, while past tense is usually maintained when describing *events* or *observations* contingent to the study itself, i.e. “... He **found** that [average time] **gets** later and later as the children **get** older until they **reach** about 20 years of age. At that age, the tendency **is** suddenly reversed, as the people in the study **started** going to bed earlier and getting up earlier.”

## Plural

Arabic has dual and plural. Dual is required in sentences where the English plural refers to two persons, such as “*Here is part of a conversation between two people who read “The Ainu and the Fox”*” (R405 Ainu). Note that in this case, where the two persons are a boy and a girl, the verb “*read*” will be in masculine dual (masculine gender being used in the Arabic grammar when referring to both males and females). If two girls had been represented in the picture, then the verb “*read*” would have to be in feminine dual.

When referring to line numbers for a quotation from the text, singular will have to be used when only one line has to be retrieved in the Arabic version (e.g. “*line 18*”), dual when the quotation has two lines (e.g. “*lines 22-23*”) and plural when it has more than two lines (e.g. “*lines 13-17*”)

Choosing between plural and dual can be tricky for Arabic translators in certain subtle contexts. In R404 Sleep, Question 10 asks the students: “*If the sets of ages in the table “Welcome to adult life” were about a different country (not Norway), would the sets of ages in the table change?*” One of the translators used the dual form for “*sets of ages*”, which assumes that only two sets of ages are shown in the table. However the table has four pairs of cells (*What the law says* and *What the body says* about *Girls* and about *Boys*), and each pair refers to two ages. Both reconcilers considered that the plural form was more appropriate in this case, since in fact the table contains more than two age groups, thus the dual form might be confusing for the student.

## Place of adverbs

Adverbs can be separated from the verb they modify in English sentences, such as in “*he had felt his opponent’s fear so strongly*” (R407 WldChp). In Arabic “*so strongly*” should be placed next to the verb “*felt*”, otherwise the adverb would be interpreted as if it referred to “*his opponent’s fear*”.

## Grammatical structure in multiple-choice items

Question 6 in R405 Ainu reads as follows:

“*These were the last words of the Ainu elder before he died*”. *What is one difference between this sentence and the rest of the text?*

- A. *It is about a different person.*
- B. *It is about a different place.*
- C. *It has a different message.*
- D. *It has a different narrator.*

Two different structures are used for the proposed answers (“*It is...*” in A and B and “*It has...*” in C and D) in a way that does not give the student any special clue about the key answer (D). However, the two structures would be difficult to maintain in Arabic, where the answers were translated into:

- A. *It is related to a different person.*
- B. *It is related to a different place.*

C. It includes a different message.

D. The narrator is different in it.

This might affect the relative attractiveness of the various answers.

## Vocabulary

The words “*adult*” and “*adulthood*” in R404 Sleep were highly confusing for the Arabic translators. The word used in Arabic (“*adult age*”) can both refer to “*access to puberty*” or to “*adulthood*” (with the more general English meaning “*when the youth is over*”). This is an interesting case when the very content of the PISA text (mentioning cultural differences in the definition of adulthood) is mirrored in the problems encountered by translators.

When used in a scientific context such as “*Professor Rönneberg estimated the average time in the middle of the sleep periods*”, the English verb “*to estimate*” would better be translated into “*to calculate*” than into the common equivalent word in Arabic, which means “*to predict*” or “*to guess*”. Similarly, in “*A new study suggests that...*” and in “*Professor Rönneberg assumes that...*” (R404 Seep), the verbs “*to suggest*” and “*to assume*” sound awkward in Arabic. The expressions commonly used in this context would be “*the study shows*” and “*Professor Rönneberg deduces*”. All three examples seem to indicate that scientific texts in Arabic would tend to suggest more “certainty” in their vocabulary than is usually the case in English scientific literature.

The word “*fable*” means “*untrue*” in literate Arabic, which would be inappropriate to translate the question stem “*Use the fable “The Ainu and the Fox” on the previous page to answer the questions that follow*” (R405 Ainu). It had to be replaced with “the story”.

Arabic has two different words for “*deer*”, depending on the geographic areas where these animals live. In R405 Ainu, since the text refers to a Japanese location, the translators should select the word that is appropriate for Asian areas.

There are also two different words for “*silence*” in Arabic, one of which is used to refer to people being quiet, while the other refers to silence in the environment. In “*He would have liked just a bit of silence after such a terrible fight*” (R407 WldChp), the latter is preferable.

In R405 Ainu the word “*Grandfather*” was translated into “*Sheikh*” (i.e. “*Old man*”, also used to refer to the leader of a community), which was accepted by one of the reconcilers, but was corrected into “*ajouz*” (“*the elder*”) by the other reconciler. In fact, adapting this word as *Sheikh*” would be questionable in PISA. First, although the Japanese fable suggests that the Ainu elder was respected and had some authority in his village, there is no specific indication that he was “*the leader of the tribe*”. Second, the fable is clearly located in a context that is meant to be unfamiliar to the vast majority of PISA students (including most of the Japanese students) through the use of foreign words (*charanke*), toponyms (*lake Shikotsu*, *Ishikari river*), and ethnic names (*Ainu*). Using the word “*Sheikh*” would be more meaningful if the entire fable was adapted to an Arabic context, which is not permitted in PISA for literary texts.

The verb “*to list*” does not exist in Arabic, although the noun “*list*” exists. Thus the sentence “*The table below lists some typical actions of a boxing champion after a victory*” needs to be adapted into “*The table below shows a list of...*”

When words are used in a religious context, appropriate terminology must be used in Arabic. In the sentence “*If the gods hear only his side of the story, they will agree to his request*” (R405 Ainu), the verb “*to agree*” was translated into “*to obey*” by one of the translators, which is inappropriate when referring to God, who can “*listen*” to a request and “*make it true*” but is certainly not supposed to “*obey*” it. Similarly, in “*We solemnly apologised to the fox god*”, the verb “*to apologise*” needs to be replaced with the equivalent verb used in religious circumstances.

The same applies in legal contexts. In “*The law says...*” (R404 Sleep), the verb “*to say*” needs to be translated in Arabic into a more appropriate word, meaning “*to dictate, to intimate*”.

### **Idiomatic expressions**

A literal translation of “*moving away from home*” (R404 Sleep) would sound too vague in Arabic. The expression had to be adapted into “*leaving the home of one’s parents*”.

“*I nodded off to sleep*” (R404 Sleep) must be replaced with an equivalent Arabic idiom, rather than translated literally.

The expression “*well-earned*” in “*They have a well-earned reputation for being late sleepers*” (R404 Sleep) proved difficult for one of the Arabic translators, who used “*non-questionable reputation*” – not quite the same meaning as in English. The second translator and the reconcilers preferred to translate into “*well-deserved*”.

In the sentence “*they are too lazy to get up in the mornings*” (R404 Sleep), the idioms “*too lazy to...*” and “*get up*” had to be adapted. The sentence became “*their laziness prevents them from waking up early*”.

The Arabic word for “*average*” has a technical meaning (“*mathematical average*”), but it can also be used with a “*common language*” meaning (“*typical*”, “*standard*”), which may be a problem when translating different occurrences of this word in a same unit, such as “*on average girls enter puberty earlier than boys*” and “*the average boy*” (R404 Sleep).

Arabic does not have a proper expression to translate “*to head upriver*” in “*the river teemed with salmon heading upriver to lay their eggs*” (R405 Ainu). The expression was adapted into a more explicit “*swimming upwards against the stream*”.

One of the Arabic translators translated “*He was feeling as miserable as death*” (R407 WldChp) into “*depression almost kills him*”. The reconcilers preferred to translate the sentence in a literal way “*he was as miserable as death*”.

### **Foreign words**

When foreign words are used in the text, such as “*charanke*” in R405 Ainu (“*The fox was making a charanke*”), the translator will need to transcribe them phonetically into Arabic letters (for example, a “*t*” will have to be added at the beginning of the word, since “*ch*” sounds “*tsh*”) and to decide whether to use masculine or feminine gender across all occurrences of this word in the whole unit. Further, as “*making a charanke*” would sound awkward in Arabic, the translators tended to paraphrase the expression into “*making a charanke complaint*” or “*singing a charanke tune*”.

## Tables and graphics

All graphics and tables (including tables presenting items and answer categories in complex multiple choice questions) need to be inverted from right to left in Arabic.

In questions based on a drawing where Speaker 1 and Speaker 2 are expressing different opinions, the dialog balloons need to be inverted too, which implies inverting the whole drawing from right to left. Note that since “*Speaker 1*” is a boy and “*Speaker 2*” is a girl, and the word “*Speaker*” must have gender marks in Arabic, masculine and feminine marks should be added to these words both in the graphic and in related questions.

## Abbreviations

Abbreviations and acronyms are relatively uncommon in Arabic. In R407 WldChp, the abbreviation “TV” was replaced with the full word “*television*”. The English abbreviations “*am*” and “*pm*” will be adapted into “*after midnight*” (“*baada mountasaf al layl*”) and “*before midnight*” (“*kabla mountasaf al layl*”). For example, in the graphic “A night’s sleep” in R404 Sleep, the text box containing the caption “*Time (am) halfway between bedtime and waking*” will need to be enlarged when replacing “*am*” with the full Arabic expression “*baada mountasaf al layl*”. It will be important that all cases where similar adaptations are used to replace an abbreviation found in the source version are accurately documented in the *National Adaptations Spreadsheets* and submitted for approval to the Consortium referee.

## Use of capital letters

Capital letters can be used in English to indicate titles, such as in “*in the journal Current Biology*” (R404 Sleep). No capital letters exist in Arabic, so this expression had to be adapted into “*in an article in the journal ‘current biology’*”.

The emphasis on the word “*NOT*” in “*What statement, NOT from his study, does Professor Rönneberg use to support his theory*” (R404 Sleep) can be maintained in Arabic by underlining rather than capitalizing it.

## The “Support your answer” issue

Like in a number of other languages, using a strict equivalent of the verb “*to support*” in the instruction “*Support your answer*” (e.g., in R404 Sleep, question 9) would result in an unfamiliar wording for Arabic students. One of the reconcilers suggested using an expression meaning “*explain argumentatively*” in all occurrences across the PISA materials. The other reconciler proposed a verb meaning “*to prop up*”, “*to underpin*”.

## English shortcuts

Certain very concise English expressions need to be translated into a more explicit wording in other languages. In “*All of his corner had jumped up into the ring*” (R407 WldChp), the Arabic translators had to add a clause to make the sentence understandable: “*All those who were in his corner had jumped into the ring*”.

Both the Arabic translators had hard times in translating “*He was aware of his face swelling and drawing tight*” (R407 WldChp). “*Drawing tight*” was translated into “*some parts of his*”.

*face stretched other parts*” by one of them, and *“his face was getting smaller”* by the other. The reconciler used the word *“stretched”*.

### **Cultural sensitivity issue**

Both the Arabic translators and the reconcilers commented that the reference to multiple gods in R405 Ainu (e.g. *“the god and goddess of the Ishikari River”*, *“the god of water and the god of mountains”*, *“We apologised to the fox god. The other gods heard about our apology”*, etc.) would appear as deeply offensive in most Arabic-speaking countries, where Islam – a monotheistic religion – is predominant. Some of them suggested using *“God”* in the singular and expressions such as *“the creator of water, of mountains, of foxes”*, etc. all across the unit. However, this adaptation would need large rephrasing of the text and it would substantially alter the cultural context of the story and the main message conveyed, which is not acceptable in PISA.

When such issues are encountered in a unit, translators should report the problem to their NPM, who is in charge to assess the suitability of the materials for the country’s students, and who can recommend that the unit be dropped from the PISA assessment in cases recognised as too sensitive. Indeed, R404 Ainu was dropped from the materials before the Field Trial phase.

## POTENTIAL LINGUISTIC PROBLEMS IN CHINESE

### Word order

Chinese grammar usually places adverbs of frequency, adverbs (adverbs of manner, adverbs of place, adverbs of time...), and all types of clauses in the beginning of Chinese sentences. This different word order creates a common translation challenge, i.e., translators try to make a sentence fluent by moving a clause to the beginning of a sentence, often dropping some of the details from the source version. For instance, “*Are there any tell-tale biological signs that youth is over and adulthood has begun?*” (R404 Sleep) was translated into “*When youth enters into adulthood, are there any biological signs?*”

In the sentence “*The transition from youth to adulthood is very much an interplay between social, psychological and physiological factors*”, the subject “*transition*” can not be placed in the beginning of a Chinese sentence, since a clause is usually placed before a noun in this language. The sentence would start in Chinese with “*Youth transfers into adulthood...*” or “*From puberty to adulthood...*”.

The word order of the sentence “*Groups of participants of different ages (e.g., 10, 12, 14 and so on) are represented by a dot on the line*” (R404 Sleep) will become quite awkward in Chinese: “*Different ages (e.g., 10, 12, 14 and so on) participants groups on the line by a dot are represented*”. This explanation may become more difficult to understand in Chinese, because the clause into parentheses must be placed between “*Different ages*” and “*participants groups*”.

Condition clauses, e.g. “*as they get older*” in the sentence “*People’s sleeping patterns change as they get older*” (R404 Sleep) are usually placed before the verb that is to be modified. And when a condition clause is long, it is usually placed in the beginning of the sentence. Consequently, the sentence becomes “*As people get older, their sleeping patterns change*” in Chinese.

Similarly, clauses are placed before the noun that is to be modified. When a nominal group contains a long relative clause in English, such as in “*What statement does Professor Rönneberg use to support his theory that changes in sleep patterns are due to biological factors?*” (R404 Sleep), translators will tend to rephrase the whole sentence, and to omit a few elements for fluency (e.g., “*changes*”, “*his*”): “*What statement does he use to support the ‘sleep patterns are due to biological factors’ theory?*”.

For this reason, long and complex English sentences including several subordinate clauses or/and several relative clauses are particularly challenging for Chinese translators. Almost the whole of the sentences in the “*World Champion*” stimulus (R407 WldChp) were considered as “hard to translate” by both translators. For sentences such as “*He said two sentences he knew by heart, the same ones he had heard and repeated a thousand times before the world championship, the same ones he was to have said with the greatest joy and that he was now saying on the verge of tears*” it is virtually impossible to find stylistic devices that convey the same rhythm as the English sentence.

Ante-position of attributes (where a semantic attribute is used as the subject of the sentence, while the semantic subject is switched as an attribute) is a common syntactic tour in English,

used to emphasise certain parts of the sentence, such as in “*One habit traditionally associated with teenagers is going to bed late*” (R404 Sleep). In order to maintain the same emphasis in Chinese, the sentence will have to be turned into “*One habit is traditionally associated with teenagers, that is going to bed late*”.

## Plural

There is no grammatical plural in Chinese. “Counter” words or other expressions need to be added in the sentence in all cases when plural is relevant for conveying the meaning of the source text. For example, in “*The Ainu and the animals lived peacefully together*” (R405 Ainu) the plural of “*animals*” can be rendered by adding after “*animals*” a noun (個) that indicates plurality. In the sentence “*They develop a habit of sleeping through the first classes of the day*” (R404 Sleep), the expression “*the first classes of the day*” will be translated into “*the first **several** class(es) of the day*”. In the expression “*This study and similar studies*” (R404 Sleep), the words “*similar studies*” need to be translated into “***other** similar study*” to indicate that there was more than one “*similar study*”.

Conversely, some determinant may be needed in cases when the word must refer to a specific singular object. Since no definite article “*the*” exist in Chinese, the sentence “*What is the purpose of the article?*” (R404 Sleep) will be translated into “*What is purpose of article?*” or, better, “*What is purpose of **this** article?*”

To indicate that “*the fox*” is the specific fox mentioned in the story, not just “*a fox*”, in “*Because the fox made a charanke*” (R405 Ainu), Chinese translators will use “***this** fox*”.

## English plurals used to convey a generic meaning

When no indication of quantity nor any determinant is used, Chinese words may be used with a *generic* meaning to translate sentences such as “*Teenagers go to bed late*”(R404 Sleep), where English and other languages use plural to indicate that *teenagers* refers to the whole category of *teenagers in general*, rather than to specific individuals.

To catch the generic meaning of the sentence “*Societies have different definitions of adulthood*” (R404 Sleep) (whose translation into Chinese would be, literally, “*Society to adulthood’s definition different*”), Chinese students will have to draw more inference from the context than English students. While they will easily understand that there is more than one *definition* (due to the word “*different*”), it will be less self-evident for them to decide whether *society* refers to *all societies in general*, each of them with a different definition of adulthood, or to a *single society*, with definitions of adulthood that may vary, e.g., across different social groups. However, Chinese students would usually assume that the word “*society*” used alone means more than one society, since otherwise it would have some determinant such as “*one*”, “*this*” or “*that society*”.

Note that generic meaning can also be conveyed in various other ways in English, e.g. by using singular with or without a definite or un-definite article, such as in: “*The Ainu did not create **the** salmon. It was the gods who created **salmon***” (R405 Ainu), or in “***a** teenager can be sleepy through the first classes of the day*”. Attention should be paid to convey the same generic meaning when translating these sentences into Chinese.

## Plural or singular? The “*Which of the following?*” issue.

The distinction between singular and plural is particularly tricky for multiple choice items where the English question contains the expression “*Which of the following ...?*”. For example, in the question “*Which of these ideas from the text is reinforced in the graph “A night’s sleep”?*” (R404 Sleep), the verb “*is*” is singular, indicating that only one of the

proposed answers is correct. Since Chinese verbs do not change their forms according to singular and plural, the sentence will have to be translated into “Which **one** of these ideas from the text is reinforced in the graph?”, making the question more explicit than in English. Conversely, when the English question suggests that more than one of the proposed answers may be selected, such as in “Which of these ideas from the text **are** reinforced in the graph?”, the Chinese translation will be more ambiguous, since the pronoun “which” can be interpreted both as a plural or a singular.

Translators must therefore be very careful when translating this type of questions. In the questions “What is one difference between this sentence and the rest of the story?” (R405 Ainu) and “Which one of the following statements gives the best summary of Professor Rönneberg’s hypothesis?” (R404 Sleep) both translators omitted the word “one”, thus the Chinese question could be interpreted as asking either for a single or for multiple answers.

## Tense

Verbs have no tense in Chinese. The common way to indicate tense is to add an adverb of time, such as “already”, “once”, “today”, “yesterday”, “last time”. As a consequence, Chinese-speaking students would have more clues that the story is situated in the past in a sentence such as “In those days there were plenty of wild deer in the mountains” (R405 Ainu) than in the sentence next to it in the same text “Whenever I wanted to eat meat, I went hunting in the mountains”.

A function word (了) can sometimes be added to indicate past tense, for example after the verb “to pass” in “As the years passed, people began to call me Grandfather” (R405 Ainu). However, its use is limited to certain verbs and to certain syntactic positions. For example, 了 can not be added between two verbs, such as “began to call”

Subtle time nuances in literary texts like “World Champion” (R407 WldChp) are often difficult to render in the Chinese translation, particularly when past, present and future tenses are used in a same passage, e.g. “He would have liked to look at him. Perhaps even to clasp him in his arms, but not the way they do it on television, clasp him as you would when you have been very afraid and you are aching and swelling all over. Just as you clasp each other when you know that the next day each of your limbs will be burning, and that for some weeks to come you will still be counting the blows”. Our Chinese reconciler was asked to comment on how the problems were solved for this specific passage. She wrote:

“For this passage, the translator added “originally” (原本) in the sentence “He would have liked to look at him” to represent the perfect tense. The whole sentence “Perhaps even to clasp him in his arms, but not the way they do it on television, clasp him as you would when you have been very afraid and you are aching and swelling all over” is rendered into the present tense. It is possible to add a function word 正在, meaning “in the process of; in the course of; in the middle of; in the act of” to indicate the progressive tense in “you are aching and swelling”. The function word 將, meaning “to be about to” can be used to express the future tense in the sentence “your limbs will be burning”. The word “still” (仍) often implies future tense in Chinese; thus the function word 將 is omitted for fluency in the sentence “you will still be counting the blows”. Sometimes the function words expressing tenses are omitted for fluency when the context provides enough clues for the tense”.

## Verbal modalities

It is tricky to find a proper word in Chinese to convey the meaning of the auxiliary verb “*may*” in a sentence like “*Girls may become adults earlier than boys*” (R404 Sleep), because “*may*” is not only used to express possibility like “*might*”, but also used to express contingency, in clauses indicating permission, condition, concession, purpose, result, etc. Translators tend either to omit “*may*” if its meaning is not explicit in the context, or to render “*may*” into the Chinese equivalent of “*might*”.

## Use of wh- words

There are two common translations for “*when?*” in Chinese, “何時” and “什麼時候”. Both convey the meaning and the ambiguity of “*when*” in English. Given the context, both are acceptable for translating “*When do we become adults?*” (R404Sleep). The reconciler used the first, which is a literary word, less used in spoken language.

Note, however, that in cases when the question refers to clock time, such as in sentences like “*When do you go to school?*”, using either of these two words would be more ambiguous for Chinese students, who might reply “*at 9 o’ clock*” or “*in the morning*”, but also “*yesterday*” (given that there is no past tense in Chinese). Therefore, when translating passages referring to clock time, a more explicit expression meaning “*at which hour?*” would be needed (幾點鐘).

Questions where the wh-word is related to the subject are easier to translate than when it refers to the object. For example, in “*Which one of the following statements gives the best summary?*” (R404 Sleep), the English word order can be respected and the sentence can start with “*which*” in Chinese. However, when “*which*” represents an object (e.g. in “*Which of these ideas does the table support?*” the subject has to be placed before “*which*”, or the sentence has to be turned into passive.

The translation is the same for “*whether*”, “*whether or not*” and “*if*”, in Chinese, e.g. in the sentence “*Whether or not one is regarded as an adult depends partly on cultural circumstances*” (R404 Sleep). In all cases when two or more of these expressions are used in the same English passage, this variation can not be respected in Chinese. For example in “*They married without knowing **whether or not** they would get along: he wasn’t sure **whether** she was able to cook, and she couldn’t tell **if** he liked to read*”, Chinese would use a same word for all three expressions in bold.

## Comparative expressions

In comparative expressions, if the second term of the comparison implies a pronoun in English, it needs to be changed into a noun in Chinese. For example, when translating “*In some cultures, one is regarded as an adult at an earlier age than in others*” (R404 Sleep), the expression “*than in others*” becomes “*than in other cultures*” in Chinese. This creates a redundancy that may be problematic for test questions related to comparative expressions. In fact, a number of PISA 2006 Science items that contained comparative expressions appeared to have DIF (Differential item functioning) in most Asian languages, and particularly in the 3 Chinese versions used in PISA 2006.

More generally, the second term of the comparison needs to keep the same structure as the first. For example, the sentence “*Sleep patterns are more related to biology than to social habits*” (R404 Sleep) will become in Chinese “*Sleep patterns are more related to biology than related to social habits*”.

If the second term was a longer group of words in English (e.g., “*social habits that people is familiar with nowadays*”), the Chinese translation will become long and difficult to read. To make the sentence fluent, some translators would change “*more*” into “*exceed*”. Another common usage in Chinese is to move a long comparative element to the beginning of a sentence. For instance, “*Compared with social habits that people are familiar with nowadays, sleep patterns are more related to biology*”. This type of rewording can have an impact on the meaning of the sentence.

## Numerals

The Chinese writing system includes both the Arabic numerals (1,2,3...) and the Chinese numerals, which has several systems:

(一、二、三...) or (壹、貳、參...) or (甲、乙、丙...).

Use of characters is roughly equivalent to writing a number in letters in English. For example, in the sentence “*In 2004, he conducted a study of 25 000 persons aged between eight and 90*” (R404 Sleep), Arabic numerals will be used for “2004” and “25 000”, but since “8” is spelled out into “*eight*” in English, it may be more equivalent to use “八” instead of “8” in Chinese.

It is interesting to note that both translators used characters to represent “1” and “2” in “*Speaker 1 / Speaker 2*” (R405 Ainu). This is a common usage in Chinese when the numbers refer to persons.

## Pronouns

When there are several identical pronouns in one sentence, Chinese translators tend to delete some of them or to replace pronouns with nouns. This is because the Chinese sentence structure does not need every precise pronoun as long as there is enough implicit reference. For example, translating “*their*” in the expression “*salmon heading upriver to lay their eggs*” (R405 Ainu) would not sound well in Chinese. The English sentence “*he asked them to record the time they went to bed and when they woke up on the days when they did not have to go to school*” (R404 Sleep) will be translated into Chinese as “*he asked the respondents (“them” replaced with a noun) to record the time they went to bed and when woke up (“they” deleted) on the days when they did not have to go to school*”.

In the sentence “*If his theory holds, it means that girls are adults when they reach 19.5 years of age and boys become men when they are 20.9 years old*” (R404 Sleep), both “*it*” and “*they*” can be deleted in Chinese. Note that there are different words in Chinese for masculine and feminine “*they*”, though grammatically it is not wrong to use masculine “*they*” to represent females.

There also are specific “*they*” for humans, animals, things, and gods. Thus in the passage “*The fish in the river and the nuts on the trees are not for the Ainu only. They should be shared with all the animals*”, the word “*they*” will be more specific than in English.

When using certain sensory verbs, such as “*to hear*” and “*to see*”, the Chinese usage tends to omit an object after the verb. The deletion of the pronoun “*it*” will not cause problems in

simple sentences like “*I listened hard, but I couldn’t hear it any more*” (R405 Ainu), but might be more problematic in sentences with complex reference chains.

In English, the pronoun “*you*” can be used as a generic pronoun referring to an unspecified person (e.g. “*somebody*” or “*a person*” or “*people*”), but in a way that is less formal than the use of the generic pronoun “*one*”, suggesting that the author is speaking either to himself (in the second person singular) or to the readers (in the second person plural). Both the Arabic and the Chinese translators were in trouble with this generic use of “*you*” in “*Just as you clasp each other when you know that the next day each of your limbs will be burning and that for some weeks to come you will still be counting the blows*” (R407 WldChp). Like many other languages, Chinese has different words for singular and plural “*you*”, therefore the translator can’t maintain entirely the ambiguity of the English “*you*”. However, the generic meaning will be conveyed by the Chinese sentence.

### **Vocabulary**

In R404 Sleep, the translation of some of the keywords that appear at different places in the stimulus and in the questions (“*youth*”, “*adulthood*”, “*puberty*”) is clearly determinant for the functioning of the unit, but it will be likely to be challenging in a number of languages, due to the cultural differences that affect the very meaning of these concepts in different societies. In Chinese, the main problem seems to be with “*youth*”, which may mean any of “*young*”, “*youthful*”, “*young people*”, “*adolescence*”, “*puberty*”, or “*youthhood*”, while “*adulthood*” appears as more equivalent to the English word than it is the case, e.g., in Arabic.

In Chinese, “*table*”, “*graph*” and “*figure*” have overlapping meanings, thus it is important to harmonize the translations for these words across the PISA materials.

In the sentence “*As the years passed, people began to call me Grandfather*” (R405 Ainu), the word “*grandfather*” may refer to “*the father of one’s father or mother*” or to any “*elder person*”. Chinese has two different words for these two meanings.

The English word “*pain*” can refer to both physical and mental pain in “*The champion knows that he will be in pain for a long time*” (R407WldChp), but there are two different words in Chinese for these two meanings. Both translators choose the word referring to *physical* pain, which can make the key of Question 4 more attractive for Chinese students.

### **Literal matches involving derived forms of a same word**

In the sentence “*The changing sleep patterns reflect basic changes in the body*” (R404 Sleep) the literal match for the word “*change*” can not be respected in Chinese, because different words are used for “*change*” depending on the context. For instance, the noun “*change*” may differ from the verb “*change*”, and the adjective “*changing*” can also differ from the adjective “*changed*”.

### **Idiomatic expressions**

In the sentence “*When I put my head back on my pillow, I heard the distant voice again*” (R405 Ainu), one of the translators rendered the expression “*put my head back on my pillow*” literally, while the other found an equivalent idiomatic expression in Chinese. The latter is fluent and is able to capture the idea, but there is no word related to “*pillow*”. If a question

was asked about what does “*put my head back on my pillow*” mean, the latter would make the item easier.

Several idiomatic expressions were used in R407 WldChp: “*for some weeks to come you will still be counting the blows*”; “*the camera was coming to lick his face for five hundred million TV viewers*”; “*he was feeling as miserable as death*”. All of them proved to be very difficult to translate into Chinese. Note that Question 4 and 5 in the PISA unit asked for the meaning of two of these expressions (“*counting the blows*” and “*as miserable as death*”). As the functioning of these items was likely to be affected by the more explicit (or less explicit) translation used, both questions were eventually discarded from the Field trial materials.

### **Answer categories**

The terms “*True/False*” and “*Correct/Incorrect*” can share the same translation in Chinese, thus translators often overlook consistency and use several translations for the same answer categories throughout and within units. It will be useful to agree from the start on a uniform translation of these four words in Chinese.

Passive voice in the answer categories “*Supported*” or “*Not supported*” would sound awkward in Chinese, so they would be translated into active voice “*supports*” and “*does not support*”. For consistency, it would be preferable to also avoid passive voice in the stem (which reads in English “*Listed below are some statements that are either supported by the graph or not supported by the graph*”, R404 Sleep), and where a passive “*supported*” would need to be constructed using two Chinese characters separated by the subject “*the graph*”.

### **Use of capital letters**

Capital letters can be used in English to indicate titles, such as in “*the journal *Current Biology**” (R404 Sleep). No capital letters exist in Chinese, therefore translators should agree on a proper and consistent notation to specify that a term like “*Current Biology*” is a journal, not a phrase.

In “*Use “World Champion” to answer the questions that follow*” (R407 WldChp), the Chinese translators had to adapt “*World Champion*” into “*the article “world champion”*” to make it clear that the expression referred to the title of the story.

In the same order of ideas, the notations for the titles of books, articles, and magazines (e.g., “ ” ) need to be harmonized. Quotation marks in the translations received were sometimes represented by [ ], sometimes by 「 」 or by 《 》 . Note that although English punctuation marks are comprehensible when they are inserted between Chinese characters, it is not appropriate to mix English punctuation marks with Chinese ones.

In the PISA source versions, capital letters are sometimes used to draw the students’ attention on certain words, such as in “*What statement, **NOT** from his study, does Professor Rönneberg use to support his theory?*” Due to lack of capitals, some other notation must be agreed by translators in order to render the emphasis put on the negation.

The English use of capitals instead of quotation marks to emphasise certain words, such as in “*decide whether each of the statements below is **True** or **False***” (R404 Sleep) or in “*people began to call me **Grandfather***” (R405 Ainu) is too peculiar to be rendered in Chinese.

## Foreign names

A common usage when a foreign name is not well known in Chinese (such as “*Till Rönneberg*” in R404 Sleep) is to translate it into Chinese characters according to its pronunciation. However, different procedures can be used. One of the translators omitted the first name, which is also a common way to make a sentence concise. Another usage is to translate both first and last name, and insert a full-point (·) in between as a separation, because it is not logical to leave space between two characters. Till Rönneberg will then become 提爾 (Till)·羅尼堡 (Rönneberg). Chinese names start with a family name (last name in English), followed by a given name (first name in English). Sometimes translators reconstruct a foreign name into a Chinese name, e.g. 羅提爾 (Rön Till, the last name being abbreviated). The translation of names needs to be harmonized in order to achieve consistency across and within the units.

When a foreign word is used in the text, such as “*charanke*” in R405 Ainu, one solution is to compose several Chinese characters to represent the pronunciation of the word, but in general the composed word will not have any meaning in Chinese. For this reason, both translators included in the text the definition provided by the English footnote. However, this kind of adaptation is not acceptable in PISA.

## Abbreviations

The time abbreviation “AM” is usually adapted into “*in the morning*” in Chinese. However, in R404 Sleep Q08, some of the time indications refer to very early hours (3am, 5am) that are usually considered as part of the night, thus they will need to be adapted into more commonly used expressions such as “3:00 in the early morning” or “3:00 in the midnight”.

## The “Support your answer” problem and other issues with instructions to students

Like in many other languages, the instructions where students are requested to find in the text some information to support their answer are very tricky to translate. One of the translators changed the English sentence “*Use information from the text to support your answer*” (R404 Sleep) into “*Quote information in the text to answer*”, which is likely to elicit simple quotations rather than explanations accompanied with examples. The second translator used “*Use the information in the text as the argument in your answer*”, which is better, but still different from the English meaning. The reconciler preferred to translate this type of sentence literally, which is understandable but not very common syntax in Chinese.

It is virtually impossible to translate into a single fluent Chinese sentence complex instructions containing several clauses, such as “*Tick “Yes” or “No” in the column next to each statement in the table below to show your answer*” (R404 Sleep). This is also the case of the instruction “*Explain your answer by referring to the story, using your own words*” (R405 Ainu). Usually, these instructions will have to be split into several simple sentences.

Direct requests such as “*Explain your answers*” may sound rude in Chinese. In a number of cases, translators tended to add “*Please*” in the beginning of the sentence. This adaptation needs to be harmonized across the materials.

## English shortcuts

In Chinese, like probably in other languages, certain very concise English expressions need to be translated into a more explicit wording. In “*All of his corner had jumped up into the ring*” (R407 WldChp), the translators had to add the word “*people*” to make the sentence understandable: “*All people of his corner had jumped into the ring*”.

Similarly, a literal translation of the sentence “*He remembered having raised his arms because that’s what is done*” (R407 WldChp) would be hard to understand in Chinese. One of the translators added a few words to specify the context (“... *because that is what is done in a contest*”), while the other translator and the reconciler replaced “*what is done*” with “*done like that*”.