

Pre-editing and the use of simplified writing for MT: an engineer's experience of operating an MT system

P.J. Pym

Perkins Engines Ltd, Peterborough, UK

INTRODUCTION

By 'simplified' we mean consistently clear, succinct and readable, yet still capable of describing technical matters which are quite complicated. Our overall objective is to improve comprehension of the English source text, particularly for those persons using English instead of their native language, and to aid translation.

I think that it should be made clear from the outset that Perkins Technical Publications Department is staffed by engineers, not linguists: MT was introduced at Perkins to deal with a particular problem.

The company was established 56 years ago in Peterborough by Frank Perkins. Recently it acquired Rolls-Royce Diesels and Gardner. It designs and manufactures diesel and other engines from 20 to 1,200 bhp, which are used in a very wide range of applications. The company has distributors in about 160 countries and 85 per cent of production is exported directly to manufacturers or distributors, or indirectly, as parts of machines produced in the UK. In 1987 production world-wide of Perkins-designed engines was 400,000 units, of which 170,000 were built at Peterborough. A high proportion of exported engines creates an obvious need for translation within the company.

The Technical Publications Department is part of the Communications Department. We produce sales catalogues, installation manuals and service publications (Figure 1):

sales catalogues enable the customer or distributor to determine the specification of the engine: these are not translated.



Figure 1. Illustration of range of publications produced by Technical Publication Department

installation manuals advise the customer how to install the engine; these are not translated at present.

service publications show operators and mechanics how to service and repair the engines; these publications are translated.

There are two main types of service publication: user's handbooks and workshop manuals. The user's handbooks cover operation and routine maintenance and are aimed at the operator. They need to be written in simplified language because we cannot be sure of the standard of literacy of the operator and because handbooks may not always be available in the user's own language. Translated versions are often required in small quantities.

Workshop manuals cover such aspects as dismantling, inspection and minor repairs and are aimed at the competent mechanic. Although the mechanic will probably be better technically qualified than the operator, it is quite likely that he or she will be no more literate than the operator. Therefore, the workshop manuals also need to be written in simplified text. Again, translated versions are often required in small quantities.

Technical Publications had been concerned about the quality of written communications for some time: their structure was not always consistent and their content was often ambiguous and full of jargon. Perkins Approved Clear English (PACE) was introduced in 1980 to simplify English publications for those using English instead of their native language and to aid translation, whether carried out by traditional or computer-assisted methods.

PACE was initially based on ILSAM (the International Language for Service and Maintenance), which in turn was based on Caterpillar Fundamental English (CFE). CFE consisted of about 800 words of basic English, plus whatever technical terms were necessary to describe the product, and was an attempt by Caterpillar to avoid the need for translation.

TRANSLATION METHOD PRE-1987

Up to 1987 the translated versions of service publications were produced by the local Perkins company or distributor, using local translators. Since Technical Publications were based at Peterborough, they could not exercise much control over the content, format or corporate imaging of the translated versions. As resources of the overseas company or distributor became more scarce, the translated versions became progressively out of date. Producing this type of publication in small quantities is often expensive.

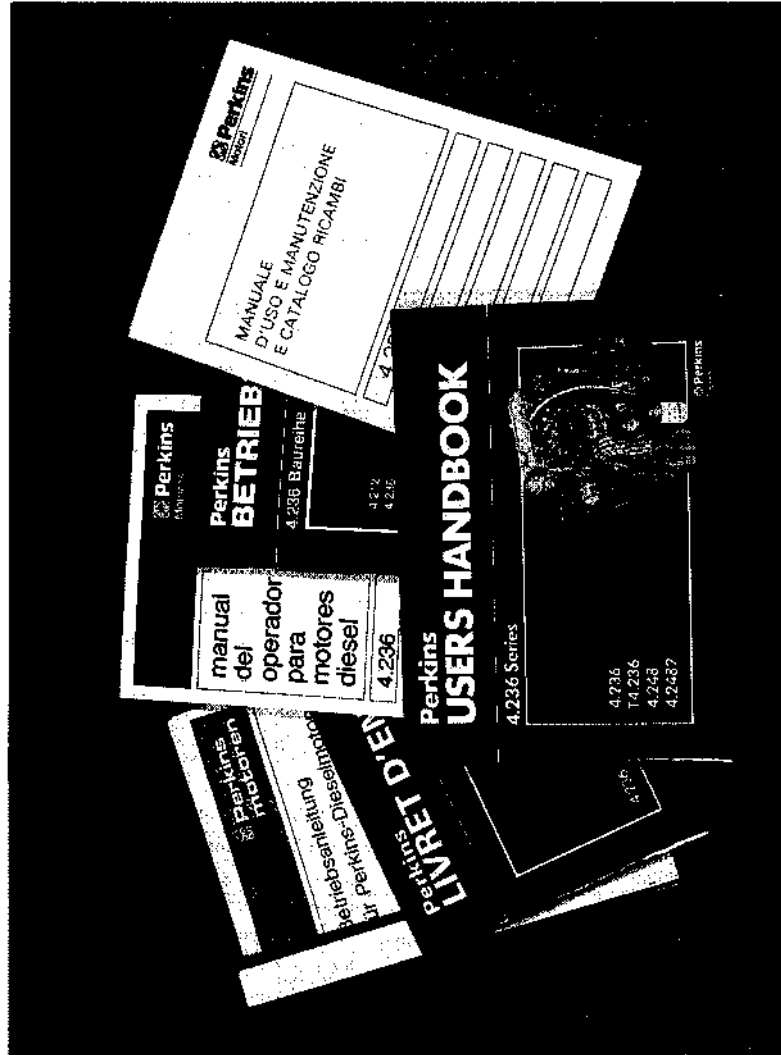


Figure 2. Mixture of styles of handbook produced for 4.236 engine.

Figure 2 shows the problem that we had up to 1987. The English handbook (bottom, centre) was A5 landscape with a fairly bold corporate image. The French and German versions were similar in format but were different in many details and the Italian and Spanish versions were quite different in every respect. For the marine version of this engine the Germans had produced an A5 portrait publication (top, left) using loose leaf sheets printed on a photocopier.

The task facing the Technical Publications Department was to produce good translations quickly, without calling on additional staff. We had to work quickly because the company was about to launch a number of new ranges of engines; we could not recruit more staff because company policy precluded the recruitment of staff at that time. The exercise also had to be self-financing by which we mean that the costs of production would have to be met from sales of the finished publications.

THE IDEAL SYSTEM

Perkins established their objectives for the ideal system in March 1984. This would comprise an electronic database incorporating a computer-aided translation system which would speed up the production of English text and translations, as well as speeding up the transmission of text to vetting parties and commercial printers. It would also reduce production costs and ensure the close control of translated versions.

The first advertisements for Weidner's MicroCat appeared in the UK in June 1984. After we had seen MicroCat demonstrated, we arranged a three month trial for the first quarter of 1985. The first PC was installed in January 1986 and three more PCs were installed in 1987.

MICROCAT TRIAL

MicroCat was chosen because it ran on a PC, used a batch system to do the translations and had its own word processor. In short, it was the best and most economical software available at that time which could meet Perkins' requirements. Tony Hartley who was lecturing in French at Bradford University at the time and had experience of machine translation, particularly the Weidner MicroCat system, acted as consultant on the project.

During the trial a handbook, written in PACE, was translated into French. We then reviewed the output with Tony Hartley and with his help devised the 'Ten Rules of Simplified Writing'. As a result of the trial, we found that we could increase the number of words in PACE, provided we did it in a controlled manner. The 'Ten Rules of Simplified Writing' were designed to satisfy two objectives:

- to make instructions clear or clearer to non-native speakers of English
- to reduce the need for post-editing.

Analysis of Perkins' earlier technical publications showed that there was a style of writing which was appropriate for this type of publication which was consistent with these objectives and was not unnatural in English.

THE PACE DICTIONARY

The golden rule for established PACE words is 'one word for one meaning'. For example, 'right' is defined as 'when facing north, right is east' and 'correct' is used to mean 'conforming to a standard, opposite of wrong'. We also try to allocate different English words to the transitive and intransitive forms of verbs, so that, if necessary, different words can be used for each sense in the target language.

Definitions are only given where they are necessary, for example, for common words or engine parts with common names such as 'part'! (Figure 3). Technical terms such as 'parent bore' are not defined because they are generally understood within the company. At the moment PACE consists of approximately 2,500 words, of which about 250 are verbs.

parent bore	(n)	
part	(n)	A service part
particle	(n)	A minute piece of matter
pass, ed	(vb)	To proceed
passage	(n)	A drilling along which a fluid moves

Figure 3. Examples from the PACE dictionary

TEN RULES OF SIMPLIFIED WRITING

The Ten rules, shown below, encapsulate points and are only meant to serve as reminders. Redundant words are those which are not necessary. Item 5 is often called the 'bracketing problem'. For example, 'A and B or C' could read '(A + B) or C', or it could read 'A + (B or C)'.

With rules 1-5 the general advice is to keep it short and simple

1. keep sentences short
2. omit redundant words

3. order the parts of the sentence logically
4. do not change constructions in mid sentence
5. take care with the logic of 'and' and 'or'.

The general advice of rules 6-10 is to make it explicit

6. avoid elliptical constructions
7. do not omit conjunctions or relatives
8. adhere to the PACE dictionary
9. avoid strings of nouns
10. do not use 'ing' unless the word appears thus in the PACE dictionary.

The practice of using groups of nouns has grown in recent years, particularly in the United States, for example, 'the cast iron timing case cover'. Europeans retain the practice of using '...of the...' so they would describe this item as 'the cast iron cover of the timing case'. By adhering to this practice not only do we ease comprehension but we also produce something which can be translated more easily. Groups of nouns, or idioms, are allowed in PACE though they usually consist of no more than two, or at the most three words. The 'ing' ending, or gerund, is avoided since it often leads to ambiguity.

Examples of how these simplified writing rules are applied

The examples given below are taken from earlier Perkins publications and illustrate various problems associated with the old form of writing. The original texts are shown at the top of each example and the PACE versions underneath. These examples, and others, are incorporated into the dictionaries that the writers use. In the first example the noun string is 'heavy duty oil bath cleaners' and the omitted conjunction is 'which'. Using PACE we have broken up the noun string by using 'of the', put in a comma and added 'which is'.

Pre-PACE

The heavy duty oil bath cleaners are usually fitted with a centrifugal pre-cleaner mounted on top of the main cleaner. . .

Using PACE

Heavy-duty air cleaners of the oil bath type are usually fitted with a centrifugal pre-cleaner, which is mounted on top of the main cleaner.

The next example shows the second part of the example above. The original sentence consists of 37 words and the phrases 'cleaned of dust and dirt' lies uncomfortably at the end of the sentence. With PACE we try to get the length of the sentence down to about 20 words, but preferably

about 15. This time we have to split the sentence into two actions, bringing the verb ‘clean completely’ forward in the second sentence. The revised version still has 29 words but that is acceptable because it is in effect a list of items. Note also the position of the adverb ‘completely’.

Pre-PACE

This should be removed and the air inlet vanes in the bottom of the assembly, the ejection slots on the side of the cone and the vanes in the outlet tube, cleaned of dust and dirt.

Using PACE

Remove the pre-cleaner. Clean completely the air inlet vanes in the bottom plate of the assembly, the ejection slots on the side of the cone and the vanes in the output tube.

The elliptical phrase in the next example is ‘shown in this section’ since we do not know which gears are illustrated – the spur gears or the helical gears? Again the word ‘which’ has been omitted. Using PACE we have re-written it slightly, inserted a comma after ‘spur gears’ and inserted ‘which are’ after ‘helical gears’.

Pre-PACE

There are a few engines fitted with spur gears instead of helical gears shown in this section.

Using PACE

Certain engines are fitted with spur gears, instead of helical gears which are shown in this section.

The next example illustrates the ‘bracketing’ problem, the incorrect use of terms according to the PACE dictionary and an unacceptable noun string. ‘Mounting’ is defined in PACE as a ‘bracket used to support a piece of equipment’ whereas in this sentence it is being used to describe the pivot fasteners. The noun string is ‘adjustment link fasteners’. When we came to re-write this sentence, we split it into two and changed the terminology. Note the repetition of the words ‘of the’ before ‘alternator’ in the revised text.

Pre-PACE

Loosen the dynamo or alternator mounting and adjustment link fasteners.

Using PACE

Loosen the pivot fasteners of the dynamo or of the alternator. Loosen also the fasteners of the adjustment link.

The last example shows a 30-word long sentence. The redundant terms are 'to a great extent' and 'it being understood'. The non-PACE term is 'carried out'. It is not clear whether 'renewing joints' refers to the operation of renewing the joints or whether it describes a type of joint known as a 'renewing joint'. Again we have split the original sentence into two. We have also done a certain amount of re-writing. Note the repetition of the word 'to' before 'renew'.

Pre-PACE

Reference to renewing joints and cleaning of joint faces has to a great extent been omitted from the text, it being understood that this will be carried out where applicable.

Using PACE

Normally the text does not include instructions to clean joint faces or to renew joints. Where they are relevant, these operations must be done.

TRANSLATION SINCE 1987

Wherever possible the English source text is prepared on the MicroCat word processor using standard copy, that is taking portions of text which already exist. The source text is fully coded in order to assist the outside printer to set the type from computer disc. The principal stages in the translation procedure are as follows:

1. prepare English source text
2. carry out vocabulary search
3. update dictionary
4. translate overnight
5. post-edit by qualified person
6. refer to authority
7. pass to commercial printer on disc.

The vocabulary search procedure lists those words in the source text which are not in the computer's dictionaries. Translations for the unfound words are obtained from our overseas customers or distributors, and the dictionaries are updated accordingly. Technical Publications staff can enter relatively simple records such as adjectives or nouns in the dictionary; verbs which require detailed linguistic knowledge are entered by the translators. The typesetting codes in the English source text are retained in the raw translation which is produced overnight, using batch processing. Experienced linguists are brought in to do the post-editing. The post-editing text is sent in hard copy form to the overseas company or distributor for checking. Finally, the polished translation is sent to the outside printer on computer disc.

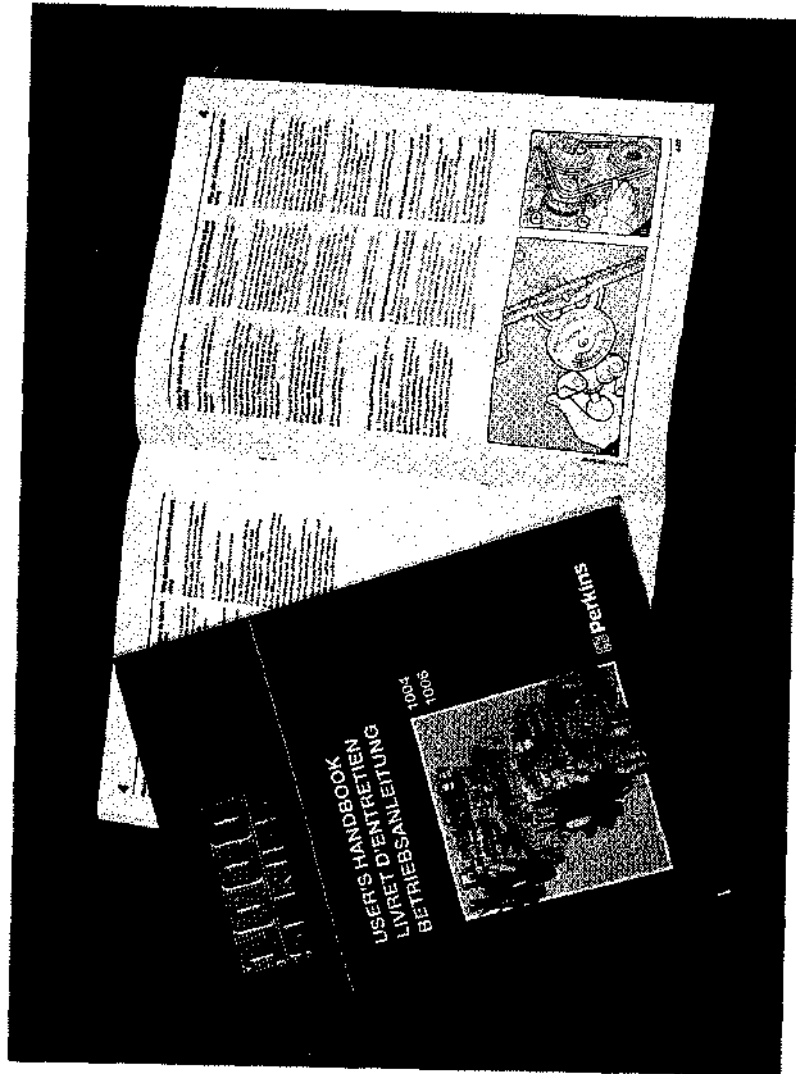


Figure 4. Illustration of multilingual handbook

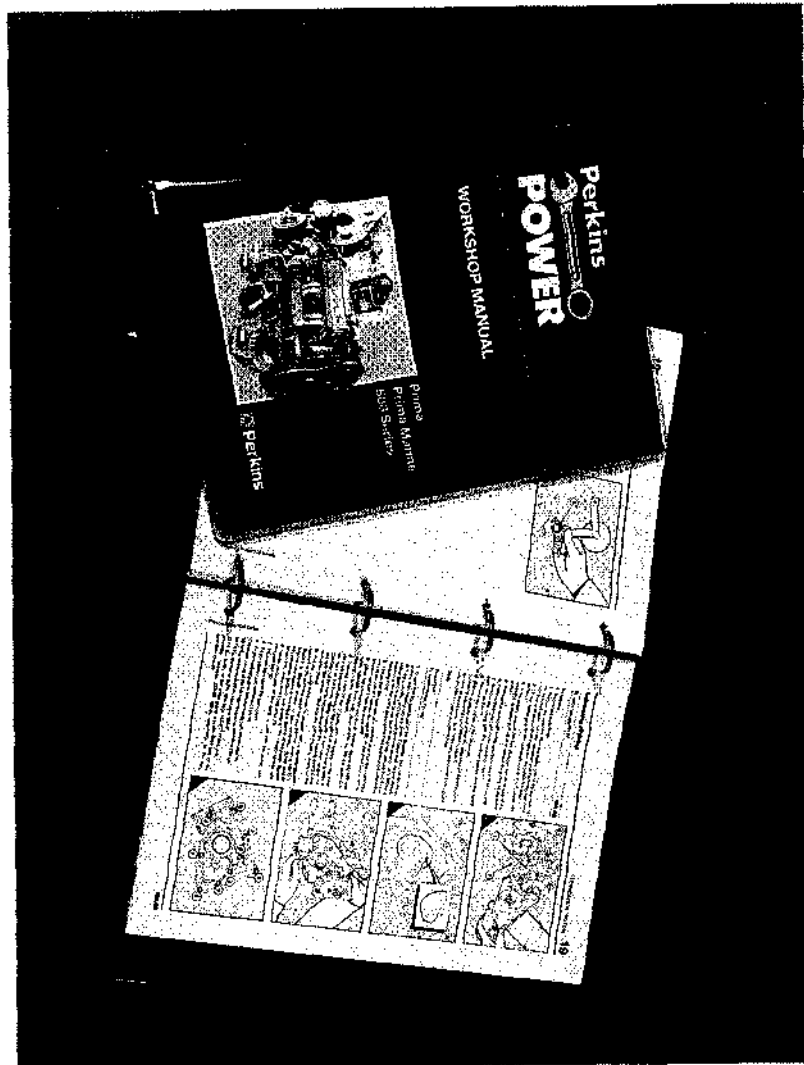


Figure 5. Illustration of workshop manual

Figure 4 shows a user's handbook for one of our new ranges of engine. The handbooks are produced in two trilingual versions: English-French-German and English-Italian-Spanish. Other trilingual versions may be produced at a later date. This format allows us to achieve the objectives of the ideal system described earlier. Essentially, it enables us to:

- maintain tight control over the content, format and corporate imaging of the publication
- keep the translations up-to-date
- reduce the number of versions in stock
- reduce costs by increasing the size of print runs.

Figure 5 shows a workshop manual for two versions of one of our new ranges of engines. The manuals are produced in single-language versions only. This is because:

- they are much larger publications than the user's handbooks, about three times the number of pages
- at the point of their use, the language is known.

The workshop manuals are produced in loose-leaf form, so that they can be amended throughout their life. Thus even though they may only be required in small quantities each year, they can still have a long shelf-life.

RESULTS ACHIEVED WITH MT

The rate of post-editing is three or four times faster than for conventional translation. For example, Perkins recently published two workshop manuals in Spanish. The first contained 63,000 words, cost £1,200 to translate (including the cost of up-dating the dictionary) and was completed in nine weeks – from beginning the translation to the delivery of the finished manual. The second contained 58,000 words, cost £900 (on the same basis) and was available 12 weeks after the English version. Previously this sort of job took at least 26 weeks and the translations would be completed some considerable time after the English version.

Text written strictly in accordance with PACE and the Ten Rules provides good source text in natural and clear English. This leads to good raw translation, which often requires very little post-editing. It would appear that not all translators like post-editing, but those that do gain satisfaction from their rate of output (between 7,000 and 9,000 words per day) and are fascinated with what the system can do.

Before we purchased this system, Technical Publications was faced with a translation load of about 200,000 words, which had to be translated into four languages over a two-year period – a potential cost of about £12,000 per language. The total cost of the software, PCs and dictionary

building will be recovered in about two years. In addition, further savings are being achieved by supplying the text on computer disc to outside printers. Overall, we estimate that translation costs have been cut by 50 to 70 per cent.

BENEFITS OF MICROCAT TO PERKINS

Using MicroCat, Perkins has been able to ensure consistent terminology and to reduce translation time as well as translation costs. Using the computerised databases, Perkins can control the source and target text at all stages of publishing. Producing translation using an MT system also ensures rigorous testing and control of the source text.

Savings of up to £4,000 in translation and origination costs are being achieved for each translated workshop manual. For the multilingual handbooks the figure is about £1,200. Without this degree of mechanisation the task would have been impossible.

Readers may be wondering whether pre-editing and the use of simplified writing for translation are worthwhile? We believe they are if the source text is to be translated into a number of languages. This is because the terminology used in the raw translations will be more consistent, thus keeping the editing to a minimum, and the source text itself will be improved, helping those people for whom we do not provide a translation.

THE FUTURE

Our objective for the future is to build on the achievements of the past. Firstly we will expand PACE, to make the writer's job easier, but secondly we will also be attempting to establish a tighter editing procedure – ideally using computer software – so that the text goes through the translation process more easily. In this way we will be able to produce English text that is consistently clear, succinct and readable, making it acceptable to native and non-native speakers of English alike. The fact that the text can be translated easily is a plus and for Perkins a most important one.

NOTES

1. ILSAM is a version of CFE and was developed by Ted White, who markets the concept in the UK and Europe. In the USA the system is marketed by John Smart of Smart Communications, New York. CFE was first used by Caterpillar Tractor Company in 1970 for writing their service publications. CFE is used by Hyster (Hyster Easy Language Program –

Help) and J. I. Case (Clear and Simple English – CASE).

Ted White can be contacted at the following address:

E. N. White, 477 Reading Road, Winnersh, Wokingham, Berkshire RG11 5HX, UK. Tel: 0734 781339

2. The Weidner MicroCat machine translation system runs on IBM PC XT or AT machines; it can also be used on some IBM clones such as the OPUS System 4, but they have to have a graphics card identical to the IBM one. MicroCat is available in the following pairs: English to French, German, Italian, Portuguese and Spanish and reverse. Each pair costs about £8,000 but discounts are available for multiple users. Weidner Translations (Europe) Ltd who supplied the software are no longer in business. At the moment the only source of information is Bravice in Japan who can be contacted by writing to:

Takehiko Yamamoto, President, Bravice International Inc., Sumitomo Ichigaya Building, 1-1 Honmura-Cho. Ichigaya, Sinjuku-Ku, Tokyo 162, Japan. Tel: (03) 235-0281, Telex: BRAVICE J 28228

AUTHOR

Peter Pym, Manager, Technical Publications, Perkins Engines Ltd, Frank Perkins Way, Eastfield, Peterborough.

APPENDIX

EXAMPLE OF POST-EDITED RAW TRANSLATION

The following three papers show some raw translation output for the Spanish workshop manual. Diacritics are converted by the multidisc reader prior to typesetting. The style and formatting codes are converted by the computer typesetting software used by outside printers. The text shows where the post-editor has made corrections to the raw translation.

<E228.1>

<h3>~~La~~ descripción general <q1>

<t3>En un motor diesel hay ^{poco} depósito de carbón y por este motivo ^{de servicio es una} ~~pequeño~~ el número de horas ~~que es~~ ^{se necesita revisar} no indicación de ~~revisar~~ cuando un ^{conjunto de culata}. Los factores que indican cuando una ^a ^{la} ^{idad} ^{de} ^{revisión} general es necesario ^{son} ~~como~~ ^{facilmente} los arranques del motor y su rendimiento general. <q1>

El conjunto de ~~la~~ ^{eleva} culata ^{una elevando} tiene dos válvulas ~~montadas~~ ^{montado con el} para cada cilindro. ~~cada~~ ^{montado con el} ~~dobles~~ ^{sencillos} ~~se~~ ^{dobles} ~~resortes~~ de válvula, según la aplicación de motor. Los resortes ^{as} ~~dobles~~ tienen bobinas amortiguador ^{que} ^{son} ^{montadas} ^{hacia} ^{la} ^{cara} ^{superior} de la culata. <q1>

Las válvulas se ^{mueven} ^{en} ~~instalan~~ ^{en} las guías fosfatadas que se pueden cambiar. La guía de la válvula de escape tiene un ^{rebajo} ~~hueco~~ ^{rebajo} contrario en el fondo y es un poco más ^a ^{larg} que la guía de válvula de admisión. <q1>

Ambos vástagos de válvula ^{elevan} ~~son~~ ^{en} ~~montados~~ ^{en} con sellos de aceite que ~~se~~ ^{montan} ^{sobre} la parte superior de las guías de válvula. <q1>

Los tipos de motor AB, AD, YB y YD ^h ^{las} ~~tiene~~ ^h insertos de asiento de válvula ~~montados~~ ^{en} la culata para ~~ambos~~ ^{ambos} ~~admisión~~ ^{admisión} y válvulas de

<E229.1>

<h3>~~La~~ cubierta de balancin <r5e><q1>

<t7>Para desmontar y montar <jr>12A-01 <r5><q1>

<h4>Para desmontar <q1>

<t3>1<s><m>Desconecte el tubo de ventilación. <q1>

2<s><m>Desmonte los casquillos roscados y las arandelas de sellar de la parte superior de la cubierta de balancin. <q1>

3<s><m>Levante ~~lejos~~ la cubierta de balancin y la junta conjuntamente con el sello de cubierta de balancin, que es montado entre la cubierta de balancin y el múltiple de inducción. <q1>

4<s><m>Cuando la cubierta de balancin es montada, los casquillos roscados se aprietan en las tuercas de los soportes de balancin. Durante el desmontaje de los casquillos roscados, es posible aflojar las tuercas de los soportes de balancin y ^{el aprieta de} estas tuercas se deben comprobar ~~para la estrechez~~ cada vez que la tapa se desmonta. <q1>

<h4>Para montar <q1>