

# Body-part nouns and local grammars \*

Jorge BAPTISTA

**Abstract:** This paper reports an ongoing study that wishes to contribute to the knowledge of the system of body-part (*Nbp*) human-related nouns in Portuguese, e.g. *cabeça* (head), *mão* (hand), and *pé* (foot). Body-parts constitute a small and rather well definable set of nouns, but they present several formal variations that render their automatic processing a non-trivial task. For this paper, I discuss the construction of a sub-lexicon of *Nbp* using local grammars for the purpose of their automatic processing in texts.

**Keywords:** body-part noun, local-grammars and electronic dictionaries, Portuguese.

**Mots clés :** nom partie-du-corps, grammaires locales, dictionnaires électroniques, Portugais.

## 1. Defining the lexicon of body-part nouns

Body-part nouns (henceforward *Nbp*) constitute a rather well definable set in the lexicon, although listing their full length in the lexicon may present some practical difficulties.

There is a rather large set of *Nbp* for non-human nouns (*N-hum*), designating the parts of plants (*branch, root, leaf*) and animals (*wing, beak, feather*). In this paper I will only consider *Nbp* that can be associated with human nouns (*Nhum*), e.g. *braço* (arm), *cabeça* (head),

---

\* I wish to thank Conceição Bravo and Ann Henshall for helping me with the English version of this paper.

---

✉ Jorge BAPTISTA, Universidade do Algarve, Unidade de Ciências Humanas e Sociais, Campus de Gambelas, P-8000-810 FARO, Portugal. Fax: +351.289.818560.  
Laboratório de Engenharia da Linguagem – Centro de Automática da Universidade Técnica de Lisboa, Av. Rovisco Pais, P-1049-100 LISBOA, Portugal.  
Fax :+351.21.8417167 e-mail: jbaptis@ualg.pt

and therefore can enter in a noun phrase with a human determinative complement: *a cabeça do João* (literally: the head of John, John's head).

Human *Nbp* can be classified in various ways. One can consider, for instance, a distinction between 'exterior' (*leg, foot, nose*) and 'interior' organs (*liver, stomach, heart*). In this paper I will only deal with exterior *Nbp*.

The list of *Nbp* can reach a significant size in scientific and technical sublanguages (consider, for example, the medical terms for the bones of the human skeleton), but at this moment I will keep to everyday lexicon.

Finally, there are many metaphorical designations of human *Nbp*. Their interpretation as *Nbp* depends on the sentences in which they appear: *Fecha as tuas asas!* (Close your wings = arms!). In this paper, I will not consider this type of expressions.

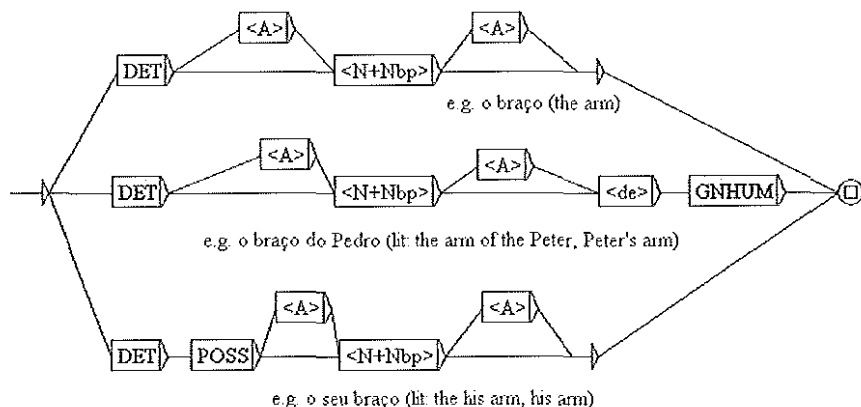
Using these rather simple, non-formal criteria, a list of about 150 human *Nbp* can be drawn, both simple (*dedo*, finger) and compound (*maçã-de-adão*, Adam's apple). The purpose of this paper is to describe these *Nbp* in an electronic dictionary in order to recognize them automatically in texts. As we will see, a simple list of *Nbp* is not enough.

## 2. Formal variation of *Nbp*

In ordinary noun phrases, *Nbp* may present different types of determiners and free modifiers. The most common cases of noun phrases whose head is a *Nbp* can be represented (in a very simplified way) by the following graph<sup>1</sup>:

---

<sup>1</sup> The graphs in this paper are finite-state automata (FSA) and finite-state transducers (FST) and they were built using the linguistic development environment software *INTEX* 4.21 (SILBERZTEIN 1993 and 2000). For an extensive overview on the use of FSA and FST in linguistic description, see ROCHE and SCHABES (eds.) 1997.



Graph 1. NP\_Nbp.grf. The most common noun-phrases with Nbp.

In this graph, gray boxes represent subgraphs: the box named **DET** calls the subgraph of determiners, **POSS** represents the set of possessive pronouns, and **GNHUM** calls for the subgraph representing human noun phrases. Categories are given inside brackets: <A> stands for adjectives and <N+Nbp> designates all nouns that were given a particular semantic information, in this case, they are *Nbp*. Other types of modifiers (relative clauses, for instance) were not taken into consideration.

The **DET** and **GNHUM** modules of this graph can be described independently from the *Nbp*. Free adjectival modifiers represented by '<A>' can be left out for the moment. This is the case of *delicado* (delicate) and many other adjectives in sentences such as:

*A Ana queimou a sua mão (E + delicada)*<sup>1</sup>  
 (Ana burned her delicate/E hand)

However, certain *Nbp* combine both among themselves and with a particular kind of modifiers that one would like to distinguish from ordinary predicative adjectives:

<sup>1</sup> Words or sequences between brackets and separated by the plus sign '+' can appear in the same position. The "E" symbol stands for the empty string. A literal translation of the examples is given to illustrate the combinatorial constraints and it is followed when necessary by a free translation in order to make the meaning clear. In the translation variants are separated by '/'.

*A Ana cortou (E + a ponta de) o dedo (E + indicador) (E + direito + esquerdo)*

(lit: The Ana cut the tip of the finger index left/right  
Ana cut the tip of the right/left index finger)

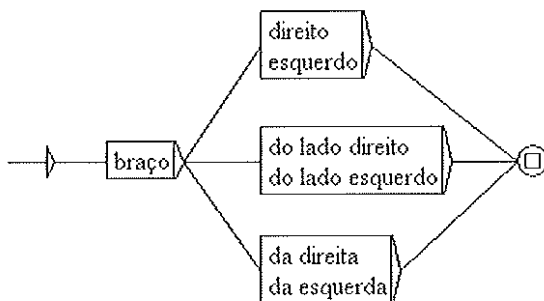
The number of combinations varies depending on the *Nbp* involved, but in some cases they can be quite numerous. It would be impossible to list them all manually. Still, as there is only a finite number of combinations for each *Nbp*, they can be described as local grammars by means of finite state automata (FSA). These FSA can then be used to adequately tag *Nbp* in texts. These combinations are to be matched onto the '<N+Nbp>' box of Graph 1, shown above.

### 2.1. Bilateral symmetry

The most important formal variation in *Nbp* modifiers derives from bilateral symmetry distinction, that is, many *Nbp* allow a modifier specifying if the *Nbp* is on left or the right side of the body. In Portuguese, this can be done in three ways:

- by adjectives *direito* (right) and *esquerdo* (left):  
*o braço (direito + esquerdo)*  
(the right/left arm)
- by a prepositional complement with noun *lado* (side) with the adjectives *direito* and *esquerdo*:  
*o braço do lado (direito + esquerdo)*  
(the right/left side arm)
- by a prepositional complement with the preposition *de* and the nouns *direita* and *esquerda*; in this case there is no noun *lado*; the two nouns are obligatory feminine and must be preceded by definite article *a*:  
*o braço da (direita + esquerda)*  
(the right/left side arm)

Usually, these three types of modifiers may all combine with a given *Nbp*. The following graph shows this formal variation:

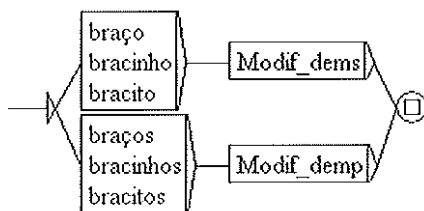


Graph 2. Illustration of bilateral symmetry opposition.

Since these three types of *Modif* appear very often with *Nbp*, subgraphs `de.grf`, `dlde.grf` and `dde.grf`, respectively, were used to represent them. These three subgraphs are called by a single graph, `Modif_de.grf`.

Gender-number agreement makes it necessary to multiply the `de.grf` subgraph by four (*ms*, *fs*, *mp*, and *fp*, where *m* = masculine, *f* = feminine, *s* = singular and *p* = plural), as well as the `Modif_de.grf`.

Finally, some *Nbp*, such as the noun *braço* (arm) allows diminutive suffixes (*bracinho*, *bracito*) and these must also be taken into consideration. The `braço_Modif.grf` that represents the formal variation associated with the *Nbp* =: *braço* (arm) will finally look like this:



Graph 3. Braço\_Modif.grf

The variation of the three *Modif* referred to above is more or less free. Usually, these modifiers appear with singular *Nbp*, but it is

possible to envisage a situation where one can use them with plural *Nbp* (see also a similar situation with *index finger* in the appendix):

?*Os braços (direitos + do lado direito + da direita) de todos os meninos apresentavam a marca da vacina*  
 (the right arms/the arms of the right side/of the right of all boys presented the mark of the vaccine)

These expressions are felt as very awkward, so in many cases we did not consider them in this paper.

## 2.2. Upper/Lower *Nbp* distinction

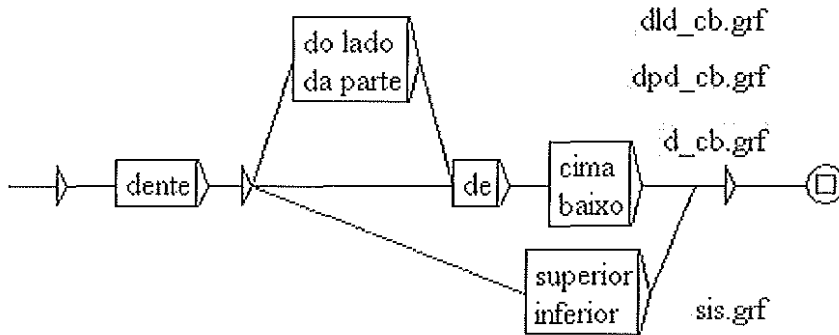
Besides the right/left opposition, there is also, but with a lesser lexical extension, the upper/lower *Nbp* distinction. This can be done at least in three ways:

- by adjectives *superior* (superior/upper) and *inferior* (inferior/lower):  
*o maxilar (superior + inferior)*  
 (the upper/lower jaw)
- by a prepositional complement *de cima* (of up, upper) or *de baixo* (of down, lower):  
*o dente de (cima + baixo),*  
 (the upper/lower tooth)
- by a prepositional complement with the noun *parte* (part)<sup>1</sup>:  
*os dentes da parte de cima (E + da boca)*  
 (lit: the teeth of the part of up, the upper teeth)

The following graph illustrates a situation of upper/lower *Nbp* opposition:

---

<sup>1</sup> The prepositional complement with the noun *lado* (side), v.g. *do lado de cima/baixo* (of the up-side / down-side) is usually less acceptable than with the noun *parte* or the two previous expressions: ?*o dente do lado (cima + baixo)* (the upper/lower tooth).



Graph 4. Illustration of the upper/lower Nbp distinction.

As for the left/right opposition, the upper/lower modifiers are called by subgraphs, whose names are shown next to the corresponding boxes in the graph above. Adjectives *superior* and *inferior* only present number inflection. *Sis.grf* (for the singular forms) and *sip.grf* (for the plural) represent these adjectives. The three types are then called by a single subgraph *cb.grf*. This has also to be doubled because of the adjectives inflectional variation, the same way as it was done for left/right opposition.

### 2.3. Classifiers

The *Nbp* =: *dente* (tooth) — but also some others *Nbp*, like *dedo* (finger) — admits a classifying adjectival modifier (and sometimes a *de N* complement), designating the different types of that *Nbp*<sup>1</sup>. These adjectives constitute a finite and rather small set (*incisivo*, *canino*, *prémolar*, *molar* and *queixal*)<sup>2</sup>. The noun *dente* (tooth) can be reduced in

<sup>1</sup> All *Nbp-classifiers* combinations are considered to be compound nouns (GROSS 1988, BAPTISTA 1995). Our point in this paper is not to determine compound nouns, but their formal variation, which is best described by means of FSA than by extensive listing of forms.

<sup>2</sup> A more technical classification of teeth uses numbers instead of adjectives to identify each tooth. A dentist, for example, would say ‘tooth 21’ for the first left incisive. In this case, the modifiers referred to above are not used. A specific-purpose graph could be built to describe this family of terms, but they were not considered in this paper.

front of these modifiers:

*o* (E + *dente*) *canino* (the eyetooth)

This makes them appear in a (superficially) nominal position, hence the fact that they are often classified in dictionaries both as adjectives and nouns. The compound *dente de siso* (wisdom tooth) also accepts all left/right and upper/lower modifiers and can appear in the reduced form *siso*. With some nouns, the left/right and upper/lower modifiers can be combined with no particular order:

*o* (E + *dente*) (*incisivo* + *canino*) (*superior direito* + *direito superior*)  
(the incisor/eyetooth upper right/right upper)

Classifiers, however, usually must be right next to the *Nbp*:

\**o* (E + *dente*) (*superior direito* + *direito superior*) (*incisivo* + *canino*)

The plural *Nbp* =: *dentes caninos* (eyeteeth, canine teeth) present restrictions on *Modif* combinations for the obvious reason that there is only one on each side:

*os* (E + *dentes*) (*incisivos* + \**caninos*) (*superiores direitos* + *direitos superiores*)  
(the incisors/eyeteeth upper right/right upper)

The following graph shows most of the combinations of *Nbp* =: *dente* and its *Modif* [next page].

The fact that some *Nbp*-classifiers allow the zeroing of the *Nbp* gives rise to a certain level of ambiguity. Some of these adjectives are unique in respect to their combination with the *Nbp*: e.g. *queixal* (*molar*) does not exist in any other combination elsewhere in the lexicon. Other words may appear both as part of *N-Adj* combinations:

*o* (E + *dedo*) *indicador* (the index finger)

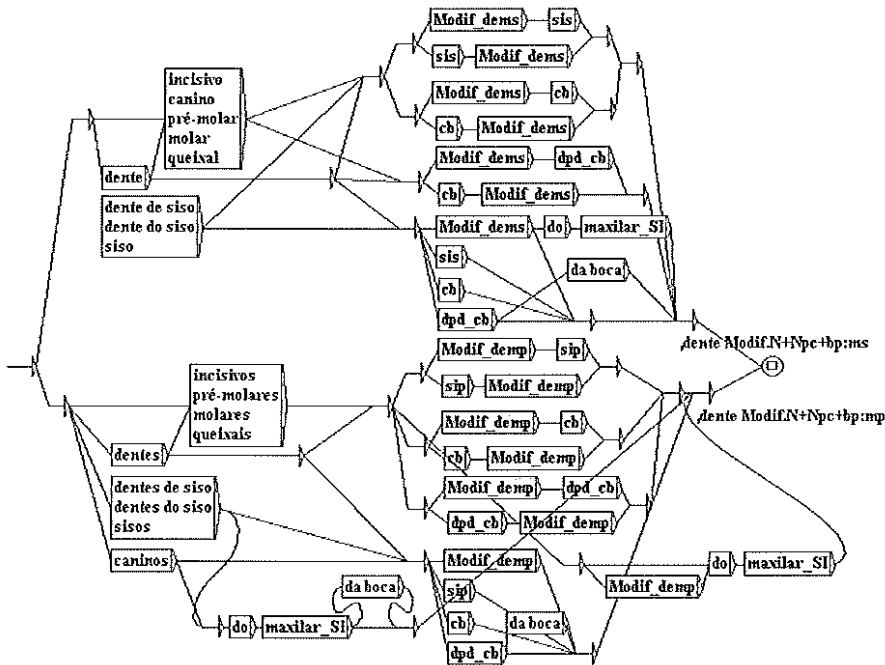
and as a simple word or part of other combinations:

*o* *indicador* (E + *económico*) (the economic index)

However, when the appropriate left/right or upper/lower modifiers are present, the reduced form of the *Nbp* (where the classifier takes its nominal position) is usually unambiguous:

*o* (E + *dedo*) *indicador* (*direito* + *esquerdo*)  
(the right/left index E/finger)





Graph 5. *Dente\_Modif.grf*.  
 Combinations of Nbp with classifiers and modifiers

In the previous case, ambiguity rises from the fact that *indicador* can be both an adjective and a simple noun, which is a common situation in Portuguese. In other cases (*incisive*), the adjective has no nominal counterpart, hence its (superficial) use in a nominal position can be identified unambiguously as the reduced form of a *N-Adj* combination.

Finally, certain classifiers can also be combined. This is the case of *dente definitivo* (permanent tooth, as opposed to *dente de leite*, milk tooth). The adjective *definitivo* can follow any *dente* + classifier combination:

*o dente (incisivo + canino + pré-molar + molar + queixal) (E + definitivo)*

but it is not allowed when the left/right or upper/lower modifiers are present. In this case, the adjective *definitivo* is felt as very awkward:

*o dente canino (superior + direito) (E + ?\*definitivo)*

For obvious reasons, *dente de siso*, which is part of the definitive dentition, also does not admit this adjective. On the other hand, *dente de leite* seems to block every other classifier:

*o dente (incisivo + canino + pré-molar + molar + queixal) (E + ?\*de leite)*

#### 2.4. Part-whole combinations

Part-whole relations between *Nbp* constitute another different situation that must be faced if one wants to find complex *Nbp* in texts:

- many *Nbp* are followed by a *de Nbp* (of *Nbp*) complement:

*a unha* (the nail)

*a unha do dedo indicador* (the nail of the index finger)

*a unha do dedo indicador da mão direita* (the nail of the index finger of the right hand)

In this last example, the last *de Nbp Adj* is equivalent to a single *Adj*:

*a unha do dedo indicador (direito = da mão direita)*

- or they can be preceded by a determinative element:

*(o canto + a ponta) da unha* (the corner/tip of the nail)

Some of these determinants can also present a 'left-right' *Modif*:

*o canto (direito + esquerdo) da unha* (the right/left corner of the nail)

and in some cases both *Nbp* can have a 'left-right' *Modif*:

*o canto (direito + esquerdo) do olho (direito + esquerdo)*

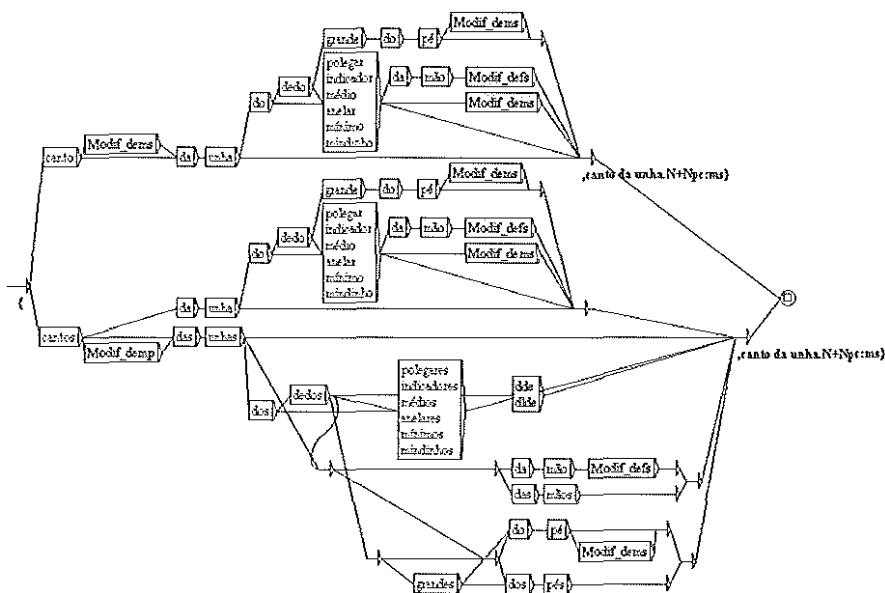
(the right/left corner of the right/left eye)

All these processes can be combined in a complex single sequence:

*o canto esquerdo da unha do dedo indicador da mão direita*

(the left corner of the nail of the index finger of the right hand)

As it is not interesting to list manually all possible combinations, one can represent them by means of a finite-state automaton. Graph 6 on the next page represents all the combinations of *canto da unha* (corner of the nail):



Graph 6. *Canto\_da\_unha.grf*. Part-whole relations between Nbp.

Some restrictions can be found in the sequences of immediately contiguous *Nbp*. While *dedo*(finger)-*mão*(hand) or *pulso*(wrist)-*mão*(hand) combinations are natural, combinations of *dedo*(finger)-*braço*(arm) or *pulso*-*braço* are not:

- <O Pedro partiu> *os dedos da mão direita* (the fingers of the right hand)
- (Peter broke) *\*os dedos do braço direito* (the fingers of the right arm)
- ?o pulso da mão direita* (the wrist of the right hand)
- \*o pulso do braço direito* (the wrist of the right arm)

The same happens with *canto da unha*:

- <O Pedro cortou>  $\left[ \begin{array}{l} o \text{ canto da unha do dedo indicador} \\ \text{(the corner of the nail of the index finger)} \\ *o \text{ canto da unha da mão direita} \\ \text{(the corner of the nail of the right hand)} \end{array} \right]$
- (Peter cut)

On the other hand, with *canto da unha* the adjective of the following *de Nbp* complement (or a further complement *de Nbp*) seems to be obligatory:

<O Pedro cortou> ?\*o canto da unha do dedo  
 o canto da unha do dedo indicador  
 \*o canto da unha do dedo indicador da mão  
 o canto da unha do dedo indicador da mão direita

at least if the first *Nbp* of the series is in the singular. If that *Nbp* is in the plural, those *Modif* may not be present, depending on the *Nbp*:

<O Pedro cortou> \*os cantos das unhas dos dedos  
 os cantos das unhas dos dedos indicadores  
 os cantos das unhas dos dedos dos pés  
 os cantos das unhas dos dedos do pé direito

### 2.5. Part-whole determinants

Some *Nbp* can also appear to the right of nominal determinants such as:

*o lado (direito + esquerdo) da cara* (the right/left side of the face)  
*a base (da coluna + do pescoço)* (the base of the spine/neck)  
*a parte (externa + interna + posterior + anterior) da coxa*  
 (the part external/internal/posterior/anterior of the thigh)

It is clear that the set of nominal determinants (as well as the modifiers they admit) varies depending on the *Nbp* they are attached to.

### 3. Conclusion

Formal variation introduced by combinations of *Nbp* with modifiers (e.g. right/left, upper/lower and classifiers) or with determinants (e.g. *canto* in *canto da unha*, corner of the nail) gives rise to an 'explosion' of combinations that will easily reach several thousand different forms. For instance, only *dente\_Modif.grf* produces over 2.000 different combinations, while *canto\_da\_unha.grf* represents about 1700 combinations.

This formal variation is of a finite nature and it is best described by means of local grammars, using finite state automata.

## References

- BAPTISTA (Jorge): 1995, *Estabelecimento e Formalização de Classes de Nomes Compostos* (MA Thesis, unpublished).
- GROSS (Gaston): 1988, "Degré de figement des noms composés", *Langages*, 90 (Paris: Larousse), p. 57-72.
- ROCHE (Emmanuel), SCHABES (Yves): 1997, *Finite State Language Processing* (Cambridge MA/ London: MIT Press/Bradford).
- SILBERZTEIN (Max): 1993, *Dictionnaires électroniques et analyse automatique de textes. Le système INTEX* (Paris: Masson).
- SILBERZTEIN (Max): 2000, *INTEX Manual* (Paris: LADL).

## Appendix: Concordance of complex Nbp sequences.

Samples extracted from a newspaper text.

<i>mão esquerda desloca-se para o lado esquerdo da anca</i>	(left side of the hip)
<i>mão direita pousa na anca do lado esquerdo</i>	(hip of the left side)
<i>as pessoas colocam um sorriso maroto no canto da boca.</i> " De resto, - <i>embora</i>	(corner of the mouth)
<i>da testa, como um indicador de perigo, o dedo indicador direito.</i> <i>Obviamente,</i>	(right index finger)
<i>a existência de uma rotura muscular na face posterior da coxa</i> <i>esquerda e vai</i>	(back of the thigh)
<i>comparando o comprimento dos indicadores da mão esquerda e da</i> <i>mão direita</i>	(right hand index fingers)
<i>por um projectil "que ficou alojado no membro inferior direito, junto</i> <i>aos testículos</i>	(right lower member)
<i>apesar de serem esquerdinos e fazerem tudo com os membros</i> <i>esquerdos</i>	(left members)
<i>por cento dos casos), afectando tanto os membros superiores como a</i> <i>cabeça</i>	(upper members)
<i>foi obrigado a desistir por causa do ombro esquerdo</i>	(left shoulder)
<i>saltaram juntos, gritando em coro "Ei! Palma da mão esquerda para</i> <i>cima,</i>	(palm of the left hand)
<i>para se alojar nas costas, já perto da pele da omoplata direita</i>	(skin of right shoulder blade)
<i>Duas no peito e duas na base do pescoço</i>	(base of the neck)

*sobre antiguidades, baixinho, óculos na ponta do nariz, mala à James*  
*Bond gasta* (tip of the nose)  
*por dentro sim, mas sempre com o rabinho do olho a espreitar a*  
*porta* (corner of the eye)  
*com quase imperceptíveis movimento <sic> da sobrancelha esquerda*  
 (left eyebrow)  
*Hoje tem o lado direito do tronco totalmente paralisado e a mão*  
*esquerda* (right side of the torso)  
*serão divulgadas as primeiras imagens da unha do dedo grande do*  
*pé esquerdo de Guterres* (big toenail of the left foot)  
*o uso do cotonete tornou obsoleta a unha do mindinho na higiene do*  
*ouvido.* (nail of the little finger)  
*Que roa até as unhas dos pés, se conseguir, mas a camisa* (toenails)