English Language and Linguistics 5.1: 131–158. © Cambridge University Press 2001.

Jespersen's formalism: problems and extensions[1]

SIGFRIDO DI GIORGI

(Received 22 February 2000; revised 15 December 2000)

This article aims to give a new, rigorous, and exhaustive presentation of Jespersen's formalism published in *Analytic Syntax* (1937). The formalism is described by means of the Backus metalinguistic formulas (Backus, 1959), which were universally adopted for the description of the syntax of computer programming languages. A modernized version of Jespersen's algorithm is proposed and illustrated with a number of significant examples.

1 Preliminaries

1.1 A historical note on Analytic Syntax (AS)

As Randolph Quirk wrote in a lecture he delivered at the Otto Jespersen Symposium, organized in Copenhagen in 1993 for the fiftieth anniversary of Jespersen's death,

> [Jespersen's] last ten years were especially, phenomenally fruitful. To mention just books alone, his *Essentials* appeared in 1933, the *Analytic Syntax* in 1937, his autobiography in 1938, Volume 5 of the *Modern English Grammar* in 1940, *Efficiency in Language Change* in 1941, Volume 6 of *MEG* in 1942, and the bulk of Volume 7 was complete by the time of his death in 1943. (Quirk, 1994: 7)

Quirk does not mention that *AS* was preceded by *Specimens of Syntactic Formulas: a Preliminary Account Dedicated to The Fourth International Congress of Linguistics* (Jespersen, 1936). The Fourth International Congress of Linguists was held in August–September 1936 in Copenhagen, and Jespersen presided over it. In his autobiography he tells us what happened:

> I had distributed to the members of the congress a small booklet called *Specimens of Syntactic Formulas*. I had interrupted my work on the large grammar to draw up a system of symbols for general grammar which in a way is a parallel to the phonetic antalphabetic designation of my youth and, like that, perhaps looks so mathematical that it frightens ordinary linguists miles away. I completed the main work itself after my return home; it was published with the title *Analytic Syntax* in January 1937 in Copenhagen and London. Whatever posterity may judge – and as yet apparently not many people have had the patience to absorb it all thoroughly – it is not too much to say that it has cost an enormous amount of work to find and analyse all those sentences from many languages and bring it all into a system. As a curiosity I can relate that a German professor, a Nazi enthusiast, to whom I sent the book, complimented me for

[1] I would like to thank the anonymous reviewers who read this article and Bas Aarts for his pertinent comments and helpful suggestions.

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

> being the first grammarian to include 'Heil Hitler!' He had probably not noticed that I also include 'A bas les dictateurs!' – but of course all the sentences I analyse are not given as expressions of my personal opinion, but solely as examples to illustrate grammatical phenomena. (Jespersen, 1995: 239)

As far as one can judge from these words, Jespersen's pioneering attempt to formalize syntax was practically ignored by the linguists of his time. The great scholar who had been born under a lucky star, who had, professionally, succeeded in all that he had tried to do, whose career had been 'an almost uninterrupted chain of successes, achieved without any of the power-struggles and the manoeuvres so often deployed in the competition for academic honours' (Bodelsen, 1964: 193), was to experience his first failure with *AS*. The book that he considered 'the crowning effort of many years' occupation with grammatical problems' (Jespersen, 1937: ix) was to be read by his fellow-students who wanted to delve into his revolutionary system of grammar, but his formalism sank into oblivion.

Only when more than thirty-two years had gone by did the community of linguists feel the need to reprint *AS*. It was in the year 1969, and the scenario of linguistics had completely changed. Linguistics had discovered in the meanwhile that a rigorous approach to grammatical theory was greatly facilitated by the use of a formalism and the advent of the computer during World War II had led to the development of recursive formulas capable of describing the syntax of programming languages. Times were ripe for a comeback of *AS*.

In his review of the reprint of *AS* James D. McCawley, who was one of the few, together with his teacher at MIT in the early 1960s, Edward Klima, who took *AS* seriously, saluted its appearance with these words:

> When I first read Jespersen's *Analytic Syntax*, it immediately assumed first place on my list of out-of print books that I wished someone would reprint : Holt and Samuel Levin deserve the thanks of the linguistic profession for making this fascinating masterpiece again generally available. (McCawley, 1970: 442)

But it was a flash in the pan and *AS* again fell into oblivion.

When McCawley reviewed *Otto Jespersen: Facets of his Life and Work* (Juul & Nielsen, 1989) he commented on the article in the volume written by W. Nelson Francis on Jespersen as grammarian:

> Francis suggests that the *AS* scheme of analysis attracted no adherents as a result of its 'omission of many features which other grammars see fit to include', such as number, person, tense, gender, aspect, and mode. Since it would be a fairly trivial matter to modify the notation so as to indicate such features explicitly, I conjecture that it was for other reasons that the scheme was, as Francis says, 'stillborn'. As more plausible reasons I nominate these: first, the notational scheme was not as well thought out as the conceptual scheme that went into it, with the result that it was disconcertingly difficult to learn to read and write it in all its baroque splendor; and second, the *AS* conceptual scheme was so far in advance of the understanding of most syntacticians of the 1930s that few of them could profit enough from the originality of Jespersen's insights to justify the investment of time and effort to learn the somewhat perverse notational system. (McCawley, 1991: 117)

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

None of these considerations is really convincing except the last, which does not account for the neglect of the formalism after 1969. Where McCawley sees baroque splendour I can only discern bright simplicity. Far from obscuring the conceptual scheme, the notational scheme clarifies it wonderfully, as every formalism always does in science. In physics a concept is perfectly clear only when it can be bounded in a formula. What Francis complains is not symbolized can hardly be said to pertain to syntax. I venture another explanation. It was one of Jespersen's greatest scientific achievements to identify syntactic relations with hierarchical relations. This concept passed to the new linguistics, but the latter saw in a linguistic structure a top–down dichotomic hierarchy between *constituents*, most of which were not recognized by Jespersen's grammar, while Jespersen saw in it a down–top nondichotomic hierarchy between *words*. For this reason the formalism was felt to be in sharp contrast with the new approach to grammatical theory, was not studied and understood in all its powerfulness, and was ultimately abandoned.

1.2 Jespersen's system of grammar

Jespersen's system of grammar is based on two fundamental ideas: the identification of syntactic relations with hierarchical relations and the distinction of *junction* and *nexus*.

In any linguistic structure, Jespersen says, it is possible to distinguish a scheme of subordination between *words* or, otherwise stated, different hierarchical levels. He calls these different levels *ranks*, and introduces the special terms of *primary, secondary*, and *tertiary* for the first three ranks, which are much more important than the others. Schemes of subordination or different hierarchical levels mean that one word is defined (or modified) by another word, which in its turn may be defined (or modified) by a third word, etc., and that only the word occupying the highest level does not depend on, and does not require the presence of, another word. In the phrase

(1) furiously (3) barking (2) dogs (1)

the numerals 3, 2, 1 denote the tertiary, the secondary and the primary, respectively, and the parallelism between the series tertiary–secondary–primary and the series adverb–adjective–noun leaps at once before our eyes. The primary is typical of nouns, the secondary is typical of adjectives and the tertiary is typical of adverbs, but this correspondence does not hold good in a great variety of cases. There may be a greater number of ranks, as for instance in the expression below:

(2) a not (5) particularly (4) well (3) constructed (2) plot (1)

But just because the series noun–adjective–adverb cannot be further carried on, there being no part of speech which bears the same relation to an adverb as an adverb does to an adjective, Jespersen grounds his grammar essentially, but not exclusively, on the first three ranks.

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

Having thus introduced the three ranks, he generalizes them to the case of sentences and so in (3) he identifies *dogs* as primary, *barked* as secondary and *furiously* as tertiary in perfect accordance with (1):

(3) Dogs barked furiously

Assigning the primary rank to the subject of a sentence was a natural consequence of the logical necessity of maintaining the same hierarchical levels for the similar words in (1) and (3). This choice was also reinforced formally by the fact that the prototypical subject of a sentence is a noun. If an intransitive verb depends hierarchically on its subject, it is obvious that a transitive verb depends hierarchically on its subject and its object, and that a ditransitive verb depends hierarchically on its subject and its direct and indirect objects. As a result of this, all the complements of a verb are assigned the primary rank.

In the sentence

(4) The rose is red

the predicative *red* is secondary, exactly as in *the red* (2) *rose* (1), but there are special cases in which a predicative may be primary. A copula and many connectives, e.g. prepositions and conjunctions, are words without a rank of their own.

In Jespersen's words, an ordinary

> verbal sentence contains a verb in one of its finite forms, and may also contain one or more than one of the following items: a subject S, a direct Object O, an indirect object [O^i], a predicative P. Besides it may comprise one or more tertiary items 3. (Jespersen, 1937: 20)
>
> S,O and [O^i] are fixed points in an ordinary sentence, placed round V [verb] as the centre . . . I have always in previous publications taken these three, S,O, and [O^i], as 'primaries'. (Jespersen, 1937: 122)

In this way all syntactic relations turn out to be hierarchical relations. This identification means that if between two words there is a syntactic relation, there must be between them a hierarchical jump or step of one unit, and vice versa. This identification is complete and without exceptions, and may be prolonged beyond the tertiary *ad libitum* until the exhaustion of all the dependencies 3 4, 4 5, etc. which are present in a sentence.

The distinction of rank, of course, applies not only to single words but also to word groups, which include all structures of more than one word which serve a unitary syntactic function or hierarchical level within a sentence. It follows that the assignment of rank depends upon the word group of which the word is a component.

Which are these word groups? Jespersen uses the terms *clause* or *subordinate clause* and *prepositional group* for some of them, but all these terms denote linguistic units which fall under one or the other of two crucial notions, namely junction and nexus, which are at the heart of his syntactic theory. In his books Jespersen, strangely enough, illustrates this fundamental distinction with many picturesque metaphors and essentially by means of semantic considerations. The simplest and

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

clearest definitions of these concepts which may be given, following in Jespersen's footsteps, are as follows:

– a junction is a structure that serves to make what we are talking about more definite and precise and the thing so defined is unique. It cannot stay by itself but requires a continuation.
– a nexus tells us something by placing two definite ideas in relation to one another through the presence of a primary/subject and either a verbal form (finite verb, infinitive, gerund or participle) or a predicative or both, or, when it assumes the form of a junction, through the presence of a primary which is a gerund and a genitive functioning as a secondary or of a primary which is an agent-substantive functioning as a subject and a secondary functioning as an object, or a primary which belongs to a special word class called by Jespersen *nexus-substantive* and one or two secondaries functioning as subject/object or as a subject and an object.

An example of junction is (1) and two examples of nexus are (3) and (4). From its definition we see that the technical term nexus includes every sentence and every subordinate clause, but also other units which it is advisable to illustrate with examples:

(5)	[He makes] her happy	a subject and a predicative
(6)	[I slept] with the window open	a prepositional subject and predicative
(7)	[I hear] the dog bark	a subject and an infinitive
(8)	[I remember] my grandfather describing this	a subject and a gerund (or participle)
(9)	Everything considered, [we decided to leave]	a subject and a participle
(10)	The doctor's coming [gave relief to the family]	a subject genitive and a gerund
(11)	The discoverer of America [was Columbus]	an agent-substantive and an object
(12)	His supporters [applauded]	an object genitive and an agent-subst.
(13)	The doctor's arrival [gave relief to the family]	a subject genitive and a nexus-subst.
(14)	The massacre of Christians by Chinese [was awful]	a nexus-subst. with object and subject
(15)	Too many cooks [spoil the broth]	a junction which is logically a nexus

The nexus-substantives are of three types: deverbal nouns like *arrival=the act of arriving*, *belief*, *sleep, fight*, etc., de-adjectival nouns like *cleverness=being clever, wisdom, rapidity*, etc., and de-nominal nouns like *friendship=being a friend, heroism*, etc. For all the verbal forms which are not finite (infinitive, gerund, participle, and agent-substantive), and also for the nexus-substantive, Jespersen uses the term *verbid*.

There is a difference in meaning between *Too many cooks* in (15) and in (16):

(16) Too many cooks are dreadfully untidy

In (16) it is possible to substitute a paraphrase with *there are*: *there are too many cooks who are dreadfully untidy*, while in (15) we must paraphrase something like: *the fact that the cooks are too many spoils the broth*. In a language there are limiting cases which cannot be discriminated formally but only semantically.

In this system the only intermediate constituents between sentence and word are

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

junctions and nexuses. The term prepositional group used by Jespersen may be considered as redundant. What else is a prepositional group if not a prepositional junction/nexus? We may have prepositional groups belonging to one of the following types:

(17) in the garden	prepositional junction (prep-junction)
(18) with the window open	prepositional nexus (prep-nexus)
(19) until he comes	prep-nexus
(20) from under the table	prep-junction with composite preposition

As an oblique (prepositional) object falls under the complement category in the same way as a direct object, so all these prepositional groups belong naturally in the junction/nexus category. The prep-nexus (19), *until* sharing the capacity of being a conjunction, obviously might also be considered a *conjunctional-nexus* (*conj-nexus*), i. e. a subordinate clause.

2 Jespersen's formalism

2.1 The Backus notation

In a previous paper (Di Giorgi, 1997) I examined the differences between Huddleston's and Jespersen's syntaxes. My overt purpose was to draw attention to the fact that the syntactic formalism presented by Jespersen in *AS* has so general a scope and so great a power of analysis as to make of it a tremendous tool for studying and comparing modern grammar systems. The greater complexity of these systems does not prevent comparison, indeed often it constitutes a spur to delve into them, bring to light many features of them that otherwise might go unnoticed, and illustrate their differences, novelties, definite pieces of progress, redundancies, incongruities, and deficiencies.

The use of the algorithm in *AS* makes Jespersen's grammar (hereafter JG) a complete formalized system, in the sense that the syntax is completely contained in the formalism, and any structure of English analysed in accordance with JG may be expressed in a formula that represents its analysis completely. The formula is a string of symbols which are very limited in number (not more than thirty, grouping and informative signs excluded) and which obey well-defined rules. Symbols and rules define the syntax fully and unambiguously.

This formalism is not only a means for a formulaic representation, pure and simple, of a linguistic structure, but has the character of a language proper, the syntax and semantics of which may be described at least partially with the same methods as those used for the description of today's programming languages. Its range of application is not limited to English but comprises a number of different languages. In *AS* there are examples from Danish, Dutch, Finnish, French, German, Greek, Italian, Latin, Portuguese, Spanish, and Swedish.

Jespersen's formalism will be described with the aid of Backus metalinguistic formulas. These allow us to furnish the syntactic description of the language

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

inherent in the formalism. They specify the syntactic formation laws of the objects belonging to the formalism or, simply, the correspondence laws between elements of the metalanguage (in our case sentences, phrases, and words of the English language) and elements of the formalism, i.e. the mapping of the metalanguage into the formalism. The Backus formulas are best explained by an example:

(21) <X>: : = S |O|<X>3|<X><Y>|

Sequences of characters enclosed in angled brackets < > represent metalinguistic variables whose values are sequences of symbols. The marks : : = and | (the latter with the logical meaning of 'or') are metalinguistic connectives. Any mark in a formula which is not a variable or a connective denotes itself. Juxtaposition of marks and/or variables signifies juxtaposition of the sequences denoted. Thus the formula above gives a recursive rule for the formation of values of the variable <X>. It indicates that <X> may have the value S or O or that, given some legitimate value of <X>, another may be formed by following it with the character 3 or by following it with some value of the variable <Y>. If the values of <Y> are given by <Y>: : = p1 (21) | (21), some values of <X> are:

(22) S,O,S3,O3,Op1(21), S(21), O(21), Op1(21) 33 . . .

In order to facilitate this study, the symbols used for distinguishing the metalinguistic variables have been chosen to be words describing the nature of the corresponding variable. The following definition will be used:

(23) <empty>: : = (i.e. the null string of symbols)

2.2 Structure and symbols

Jespersen's formalism is built up using the following notions and symbols:

(24) (a) <structure>: : = <function>|<article>|<form>|<intermediate form>|<delimiter>
(b) <intermediate form>: : =<junction>|<nexus>
(c) <delimiter>: : = <bracket>|<separator>|<declarator>
(25) (a) <function>: : =S|O|O^i|P|1|2|3|4|5| <etc.>
(b) <article>: : =
(c) <form>: : = V|v|I|G|X|Y|Z|p|pp|&|3^c|3^{cc}|3/s|s|o|o^i|+
(d) <bracket>: : = (|) | [|] | { | } | < | >
(e) <separator>: : = – |+ | /
(f) <declarator>: : = ? | $!^r$ | !! | ! | § | | ½ |

The symbols of the formalism are listed from (25a) to (25f). Each symbol on the right of the mark : : = in the formulas (25) has a specific meaning.

Functions:

S : Subject
O : Object

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

O^i	:	Indirect object
P	:	Predicative
1	:	Primary
2	:	Secondary
3	:	Tertiary
4	:	Quaternary
5	:	Quinary, etc . . .

Forms:

V	:	Finite verb
v	:	Auxiliary verb
I	:	To-infinitive or bare-infinitive
G	:	Gerund
X	:	Nexus-substantive
Y	:	Agent-substantive or participle
Z	:	Unanalysable sentence
p	:	Preposition
pp	:	Composite preposition
&	:	Coordinating conjunction
3^c	:	Subordinating conjunction
3^{cc}	:	Composite subordinating conjunction
3/s	:	Existential 'there'
s	:	Subject 'it' in extraposition
o	:	Object 'it' in extraposition
o^i	:	Indirect object 'it' in extraposition
+	:	morpheme or part of word

Unfortunately, it is impossible to describe the syntax of the junction and nexus with the Backus formulas, English not being an artificial language but a very complex natural language.We can only give their definitions, which can now be formulated as follows:

– a junction is a hierarchical scale of words, not necessarily contiguous, equal hierarchical levels being possibly repeated, with no V, no verbid I, G, X or Y, and no P.
– a nexus is a hierarchical scale of words, not necessarily contiguous, equal hierarchical levels being possibly repeated, with a V, I, G, X, Y, or P. With X and sometimes with G and Y, the nexus is only implicit and represented as a junction.[2]

The following are elementary examples of junctions and nexuses:

(26) He gave her a ring SVO^iO (*AS*: 6)
Terribly bad weather 321 (*AS*: 7)
You walk far too rapidly SV543 (*AS*: 32)
He will learn it soon enough SVO34 (*AS*: 32)
He is angry SVP (*AS*: 6)
Will he come vSV (*AS*: 6)

[2] But for P, which projects a semantic notion, these definitions would be entirely formal.

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

The norm is that a sentence is mapped into a formula word for word and every word is represented either by a symbol of function or by a symbol of form. The most important exception is the article, which is represented by the null string, i.e. it is not represented. Other exceptions are such combinations as *has taken, have been taking, can take* etc., *to*-infinitives such as *to take*, composite prepositions such as *because of* or *on account of*, and composite conjunctions such as *as long as* or *as soon as*, which are reckoned as one unit and therefore represented only by one symbol, respectively by V, I, pp and 3^{cc}. I is used by Jespersen also to represent bare-infinitives. The auxiliary is regarded only when it is separated from the main verb. Jespersen used the symbol R for a function which he called recipient, i.e. for 'something that is governed by some word that is not a verb or a verbid (I, G, X, or Y) in about the same way as an object is governed by any of these' (Jespersen, 1937: 33). R has been left out because it may be replaced without any harm by O, when it is a complement of an adjective as in *The ring is worth 5 pounds* SVP(2R(21)) (*AS*: 33), or by O^i where it is a complement of an adjective as in *Il lui était impossible de répondre* sRVPS(I) (*AS*: 34). The symbol W for the complex verbal phrases such as *waited on* in the sentence *She waited on us*, rendered by Jespersen as SWO, has been suppressed because it constitutes a further exception to the norm of one symbol for one word. This can be avoided by using the two symbols Vp, SWO changing into SVpO.

Brackets:

( ) Particulars serving to explain the item immediately preceding
[ ] Extraposition or apposition
{ } Two or more items expressed in the same form
< > Speaker's aside

The following are examples in which all the symbols already introduced are used at least once:

(27) To live is to suffer S(I)VP(I) (*AS*: 48)
Browning, the poet 1[1] (*AS*: 7)
Amo puellam {SV} O (*AS*: 7)
He fairly screamed S<3>V (*AS*: 7)
Complimenting is lying S(G)VP(G) (*AS*: 55)
The control of mind over matter XpSpO (*AS*: 57)
The discoverer of America YpO/Y2(pO) (*AS*: 59)
He stopped on account of the noise SVpp1 (*AS*: 32)
Yes ! Z! (*AS*: 80)
O, these women! Z1(21)! (*AS*: 80)
They fought tooth and nail SV3 (1 & 1) (*AS*: 31)
He ran away because he was afraid SV33(3^cSVP) (*AS*: 69)
She laughed as if she were mad SV3(3^{cc} SVP) (*AS*: 69)
There are many churches there 3/sVS(21)3 (*AS*: 33)
It is a pleasure to see you sVPS(IO)[3] (*AS*: 6)
We have it in our power to do great harm SVop1($S^2$1)O(IO_2(21))[4] (*AS*: 92)

[3] The pronoun *it* is represented by s as a 'lesser' subject.
[4] The pronoun *it* is represented by o as a 'lesser' object.

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

He never gave it a thought that Bolshies are human beings
S3^{n}Vo^{i}OO^{i}(3^{c}S$_{1}$VP(21))[5]

Separators:

– Compound of the ordinary type
+ Compound of adjective and substantive
/ Alternative analysis (the logical 'or')

Examples:

(28) Finger-ring 2 – 1 (*AS*: 8)
Holiday 2+1 (*AS*: 8)
He is in the garden SVp1/3(p1)
A blue-eyed girl 2(21+)1 (*AS*: 8)

The symbol + is used by Jespersen for two different purposes: as a separator in compounds and as a form, more precisely as a morpheme added to a word. In the above example *blue-eyed* is interpreted as a junction *blue-eye* 21 transformed into *blue-eyed* by the morpheme *-ed* and consequently represented by 21+.

Declarators:

? Question
!^{r} Request
!! Wish
! Exclamation
. . . Related distant words
§ Referring to a whole idea
½ Preceding split S and O

These symbols do not represent words but give additional information. The signs ?, !^{r}, !!, ! denote the interrogative, imperative, optative and exclamative sentences, respectively.

Examples:

(29) Who says that? S?VO (*AS*: 7)
Is that true? VSP? (*AS*: 8)
Come now! {SV}3!^{r} (*AS*: 8)
God bless the king! SVO!! (*AS*: 8)
How dreadful! 3!P (*AS*: 41)
What did they talk of? O*?vSVp* (*AS*: 26)
I saw her, however SVO[§3]
This was considered a lie ½SV½S(P) (*AS*: 46)

Many symbols may be affected by subscripts and superscripts which give them additional meanings:

S : first subject
S$_{1}$: second subject in the same sentence
S$_{2}$: third subject, etc . . .

[5] The pronoun *it* is represented by o^{i} as a 'lesser' indirect object.

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

O : first object
O_1 : second object in the same sentence
O_2 : third object, etc . . .
V : first verb
V_1 : second verb in the same sentence, etc . . .

Jespersen used the series 2,3, etc. instead of 1,2, etc. These subscripts are regularly applied to S or O but only occasionally, when really necessary, to V or other symbols.

Superscripts:

S^o : understood subject
O^r : object of result
P^r : predicative of result
1^2 : primary changed into secondary (or genitive)
S^2 : subject genitive
O^2 : object genitive
S^c : subject relative pronoun
O^c : object relative pronoun
1^c : relative pronoun primary of a junction
S^∞ : generic person

Examples:

(30) I consider this a lie SVO(S_1P) (*AS*: 6: with S_2 instead of S_1)
He wants to see her SVO(IO_1) (*AS*: 48: with O_2 instead of O_1)
He opens and shuts the door SV(V_1&V_2)O (*AS*: 21: with V and V_2 in brackets)
Live and let live V&VO^r(S^oI)!^{r} (*AS*: 44)
He makes a hole SVO^r (*AS*: 21)
He gets angry SVP^r (*AS*: 28)
He takes his hat and stick SVO($S^2$1)&O_1(S^{2o}1)/SVO($S^2$1(1&1_1)) (*AS*: 21)
His supporters O^2Y (*AS*: 59)
Jane's was a happy life S(1^2*) VP (2 1*) (*AS*: 26)
The man who killed Jaurès was not punished S(12(S^cVO))V^{bn}(*AS*: 65)[6]
The man whom he killed was Jaurès S(12($O^c$$S_1$V))VP (*AS*: 65)
I dislike the way in which he treats the children SVO(12(p1^cS_1VO)) (*AS*: 65)
To see is to believe S($IS^{o\ \infty}$) VP (I_1 $S^{o\ \infty}$) (*AS*: 128)

Other superscripts:

V^a : active (with reference to form or meaning) finite verb
Y^a : active agent-substantive or participle
S^a : what is S in the active transform
O^a : what is O in the active transform
O^{ia} : what is O^i in the active transform
V^b : passive (with reference to form or meaning) finite verb
Y^b : passive agent-substantive or participle
V^n : negative finite verb
3^n : negation 'not'

[6] See immediately below for the meaning of the superscripts in V^{bn}.

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

Examples:

(31) He was beaten by John $SV^{b}pS^{a}$ (*AS*: 26)
A reward was offered the butler $S(O^{a})V^{b}O^{i}$ (*AS*: 26)
The butler was offered a reward $S(O^{ia})V^{b}O$ (*AS*: 26)
The bed was not slept in $S*V^{nb}p*$(*AS*: 27)
The book sells well $SV^{a/b}3$ (*AS*: 27)
An admired employer $2(Y^{b})Y^{a}$ (*AS*: 60)
An admiring employee $2(Y^{a})Y^{b}$ (*AS*: 60)
A learned man $2(Y^{a/b})1$ (*AS*: 60)
This is not good $SV^{n}P(2)$ (*AS*: 28)
All that glisters is not gold $S(1*2(3^{c}V))V3^{n}*P$ (*AS*: 66)

Other superscripts and subscripts:

2^{m} : secondary of an unclassifiable junction
2^{q} : quantifier
S_{x} (read 'cross', not the letter) : subject in reciprocal relation to another item

Examples:

(32) A general paralytic $2^{m}1$ (*AS*: 12: 'formed from the regular *general paralysis*')
Many friends, a few friends, some friends, four friends $2^{q}1$ (*AS*: 15)
Very many friends $32^{q}1$ (*AS*: 15)
A great many friends $2(21)^{q}1$ (*AS*: 15)
They hate one another $SVO(S_{x})$ (*AS*: 6)

2.3 Rules

The ranks 1, 2, 3 – that is, the syntactic functions that Jespersen calls primary, secondary, tertiary – are expressed in the formalism in a different way depending on whether the forms which fill them are parts of a junction or of a nexus. Within a junction the ranks are represented by the numerals 1, 2, 3, while within a nexus, a term which embraces, among other forms, the ordinary independent sentence,

> the primary or primaries are indicated by the more specific symbols of syntactic functions S, O, O^{i}, pO, pO^{i}, the secondary or secondaries by the more specific symbols P (but in some special cases P may be primary . . .) and by the symbols of constituent V, I, Y, G and the tertiary or tertiaries by the numerals 3 followed by their description or simply by their description. In Jespersen's formalism the numerals 1, 2, 3, which typically express ranks, i.e. syntactic functions, are also extensively used to express the constituents N [or pronoun], Adj, Adv and, vice versa, symbols of constituents like V, I, . . ., p1 are also representative of the ranks implied for them by Jespersen's theory. All this results in a great simplification of formulas without any loss of accuracy. (Di Giorgi, 1997: 22)

Let me illustrate these points with examples:

(33) (a) They fought tooth and nail SV3(1&1) (*AS*: 31)
(b) This is not good $SV^{n}P(2)$ (*AS*: 28)
(c) The rain is over SVP(3) (*AS*: 29)
(d) The life to come 12(I) (*AS*: 49)

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

(e) I found John in the garden SVO3(p1) (*AS*: 32)
(f) To see you is a great pleasure S(IO)VP(21) (*AS*: 48)

In (33a) *They* is represented by S, subject/primary of the nexus SV3(1&1), the entire sentence, *fought* by V, secondary of the nexus, and *tooth and nail* by 3, tertiary of the nexus described by (1&1) as two coordinate nouns.

In (33b) *This* is represented by S, subject/primary of the sentence, *is not* by V^n, the verb being a copula and as such with no rank, *good* by P, a predicative/secondary described by (2) as an adjective.

In (33c) 3 qualifies the form *over* as an adverb because functionally it is P, i.e. secondary.

In (33d) *The life* is represented by 1, primary of the junction *The life to come*, and *to come* by 2, secondary of the junction described by (I) as an infinitive.

In (33e) *I* is represented by S, subject/primary of the sentence, *found* by V, secondary, *John* by O, object/primary, and *in the garden* by 3(p1), the nexus-tertiary of the sentence in which '1 qualifies the form *garden* as a noun, because if, in the prepositional group *in the garden* p1, which is a form, something has a function, this is p and not 1' (Di Giorgi, 1997: 31).

In (33f) *To see you* is represented by S, subject/primary of the sentence described by (IO) as a nexus of which I is the secondary and O is an object/primary, *is* is represented by V, a verb with no rank, and a *great pleasure* by P, predicative/secondary of the sentence described by (21) as a junction in which *great* is secondary, ambivalently (see below) adjective and *pleasure* is primary, ambivalently noun.

From what precedes we see that many symbols of forms also have an implicit value as nexus-ranks (usually secondaries), many symbols of functions also have an implicit value as nexus-ranks and the junction-ranks 1 2 3 . . . have an implicit value as forms, namely noun, adjective, and adverb. The nexus-tertiary not being absorbed by other symbols, as happens for the primary and secondary, the number 3 is used for it.

The presence of forms in Jespersen's formulas is neither coincidental nor negligible. On the contrary, it is highly significant. A Jespersen formula is far from being a purely functional notation: all parts of speech are explicitly or implicitly described except the pronoun (if need be, the numeral 1 could be used for it), but the three symbols s o o^i are exclusively used for *it* when performing special syntactic functions, and many symbols are ambivalent in signifying forms and functions or more than one function. It is this ambivalence that simplifies the algorithm enormously, and makes it so dense and pregnant.

In a Jespersen formula *three levels of representation* are easily recognizable:

– *The verbatim representation*. It is usually a one-to-one correspondence between words and symbols, but there are cases in which one word is represented by two symbols. When, in the one-to-one correspondence, the symbol used to represent the word is a symbol of form (V, I, G, etc . . .), this symbol also assumes a functional meaning, as required by the theory which underlies the algorithm, and

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

in this way its meaning becomes ambivalent. When there are two symbols for one word as in (33b, c, d), where *good* is represented by P and 2, *over* by P and 3, and *to come* by 2 and I, the first symbol (P, P, 2 in these cases) describes the function and the other (2, 3, I) describes the form (adjective, adverb, infinitive). The examples (33b, c, d) guarantee the ambivalent meaning also of the functional symbols of rank 1, 2 and 3.

– *The intermediate-level representation*. This is a one-to-two correspondence between word-groups and symbolic notations. Every word-group is represented by one symbol and by one combination of symbols closed in round brackets. The symbol describes the function of the word-group and the combination of symbols describes the form. In (33e, d) we have

  in the garden 3(p1) where 3 is the function (nexus-tertiary) and (p1) is the form (PP)
  To see you S(IO) where S is the function and (IO) is the form (nexus)
  a great pleasure P(21) where P is the function and (21) is the form (junction)

– *The sentence-level representation*. This is a one-to-one correspondence between the entire sentence and the formula, which represents the form of the sentence. If the sentence is well-formed (grammatical) and significant in some context, the formula describes one of the innumerable sentence-forms belonging to the language.

An analysis is defective if the rank of an item is left undecided and for this reason Jespersen establishes the convention that a p1, being 'one of the most frequent forms of tertiaries' (Jespersen, 1937: 32), must be considered a tertiary if not expressly stated otherwise, and uses p also for assigning the tertiary rank to a *to*-infinitive (e.g. a purpose infinitive), as in the example

(34) The motor car has come to stay SVpI (*AS*: 53)

He describes *to stay* by pI, and represents *to* twice, because *to* is usually absorbed in the symbol I, except obviously for the bare-infinitives. The process of sentence analysis

> is regulated by the identification of the ranks of the words or word-groups appearing in the structure being analysed . . . [In] a nexus, there are usually three ranks at most along which to distribute the words or word-groups, and each rank may appear more than once, i.e. there may be more than one primary or secondary or tertiary in it . . . [In] a junction the number of ranks may be more than three, i.e. words may appear which are quaternaries (4) or quinaries (5) and so forth . . . The rank is pinned down by assigning the appropriate symbol of Jespersen's formalism to the words or word-groups. If the ranks so found are filled by single words the process is terminated; if they are filled by combinations of words, i.e. usually new nexuses or junctions, the symbols chosen for them are followed by a pair of brackets ( ) and inside the brackets the process is applied recursively with the same criteria. The process terminates when all the words, except articles, are examined and the result is a formula which describes the grammatical construction exhaustively and in detail. (Di Giorgi, 1997: 31)

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

2.4 *Review of the forms appearing in* AS

A careful and thorough screening of *AS* has singled out the following cases of forms appearing in *AS*'s formulas which may raise some doubt:

(35) Pre-war prices 2(p1)1 (*AS*: 10)
An out-of-work joiner 2(pp1)1 (*AS*: 10)

Here *pre-war* and *out-of-work* are not prepositional groups or junctions formed freely by different words; they are compound words which are found in dictionaries and usually used as secondaries, and ought to be symbolized by 2(p-1) and 2(pp-1).

(36) It is a long way to Tipperary sVS(21p1) (*AS*: 27)

The content of the brackets is a junction, but the rank of *to Tipperary* p1 is not a tertiary according to Jespersen's convention for prepositional groups, a secondary which goes with the primary *way*. In the second alternative given by Jespersen in *AS*, namely SVP(21)p1, p1 is certainly a tertiary which qualifies the secondary P(21) *a long way*. I do not like these formulas because the first includes nearly all predication in the subject (*a long way to Tipperary is*) and the second introduces an unnecessary unspecified subject (S *is a long way to Tipperary*). The solution sVP(21)S(p1) is preferable where *to Tipperary* is primary, the extraposed subject of the sentence *To Tipperary is a long way* S(p1)VP(21).

(37) We talked face to face SV3(1p1) (*AS*: 31)

The combination *face to face* may be interpreted as a junction 12(p1), something which may be paraphrased by *face which is to face*, or as a nexus $SY^{o}3(p1)$ with Y^{o}=being, that is to say something paraphrased by *face being to face* in which *to face* is a tertiary place relation. This junction/nexus is in relation with the verb *talked* as a tertiary and with the subject as a predicative. For a form sharing more than one syntactical function see section 3.2.

(38) (a) An old man with one foot in the grave 212 (p 2^{q}1p1) (*AS*: 10)
(b) From nine till twelve is three long hours S(p1p1)VP(2^{q}21) (*AS*: 22)

In (38a) *with one foot in the grave* is either a junction or, as in (37), a prep-nexus described by p1(S(2^{q}1)Y^{o}3(p1)) with Y^{o}=being. In (38b) *from nine till twelve* may be considered as a constituent formed by a discontinuous composite preposition *from . . . till . . .* followed by its discontinuous object *nine . . . twelve*, that is to say a *discontinuous prep-junction* represented by p*1p*1. Another interpretation might be to consider one of the prep-junctions *from nine*, *till twelve* subordinate to the other (but which one is the subordinating and which one is the subordinated?), and symbolize the whole by 1(p1)2(p1) or 2(p1)1(p1), i.e. a junction having two prep-junctions as primary and secondary.

(39) (a) He lived as a saint SVpP (*AS*: 29)
(b) They were changed from men to swine $SV^{b}pPpP^{r}$ (*AS*: 30)

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

The first formula is a very dense one and must be deciphered as follows: *as a saint* is a prepositional group which qualifies *lived* as a tertiary, and its regimen *a saint* functions as predicative of the subject. A more explicative formula would be SV3(p(P(1)) with the alternative SV3(P(p1)) in which the predicative is *as a saint*. In (39b) Jespersen's formula is unsatisfactory because it says nothing about whether *from men to swine* must be considered as a constituent, and also nothing about its relation with the verb. In my view *changed from . . . to . . .* are all words occupying the verbal/secondary position (the prepositions are selected by the verb and form a whole with the verb) and *men* and *swine* are objects of the verb through a preposition and predicatives of the subject. Hence the formula $SV^{b}pO(P(1))pO_{1}(P^{r}(1))$ where the numerals describe *men* and *swine* as nouns.

(40) She presided with a grace all her own SVp1[$3S^{2}2$] (*AS*: 14)

How should we interpret the material contained in the brackets? It appears to be a junction which is deprived of its primary (*grace*). Whenever we have an incomplete junction which is saturated (completed) by a word appearing elsewhere in the sentence, we may call this junction a *quasi-junction*.

(41) That is all right SVP(32) (*AS*: 30)

Here '(32)' describes the ranks (and ambivalently the forms) of words which are part of a nexus and what is lacking is the primary/subject of the nexus, which is outside the brackets. Such an incomplete nexus may be called a *quasi-nexus.*

(42) Brown, never at a loss for an answer, replied S[3^{n}p1pO]V (*AS*: 14)

The quasi-nexus in brackets has *at a loss*, p1 (equivalent to a V), as secondary, and the prepositional object *for an answer* as a primary, and is saturated by *Brown*. A more detailed formula would be S[3^{n}2(p1)pO]V.

(43) Her temperature had dropped to normal S(1^{2}1)VpP^{r} (*AS*: 30)

The tertiary *to normal* pP^{r} is a *prep-quasi-junction*, whose regimen is a predicative of result of the subject, saturated by *temperature*. A more accurate formula would be S(1^{2}1)V3(pP^{r}(2)).

(44) The apple is far from ripe SVP(3p2) (*AS*: 29)

Far from ripe (3p2) is a quasi-nexus of which p2 is a secondary having the form of a prep-quasi-junction whose regimen *ripe* 2 is saturated by *apple*. The quasi-nexus, instead, is completed by its subject *The apple* S. It is better to write S(21)VP(32(p2/p1(2))) where 2 in S(21) describes the article.

(45) Parliament breaking up gave the officials a good excuse for doing nothing
S(S_{1}G3)VO^{i}O(21p1(GO)) (*AS*: 56)

What is the rank of p1? Indeed, what is (21 p1(GO))? It seems to be a junction but, as a matter of fact, is a quasi-nexus because *excuse* is a nexus-substantive X having *the officials* O^{i} as understood subject and *for doing nothing* as object. Thus the

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

formula is to be changed into 2XpO(GO_1)) where GO_1 is also a quasi-nexus completed by *the officials* as its subject.

(46) He comes here every now and then SV33(21(3&3)) (*AS*: 31)
He despised the Church, whether high or low SVO[&2&2] (*AS*: 36)

In (46) the contents of the brackets (3&3) and [&2&2] are *coordinate equal parts of speech*. The first & before 2 is lacking in AS. These examples are different from the sentence

(47) He arrived home late SV3(33)

where the brackets contain a *string of unconnected equal parts of speech*, which are independent one-word constituents separately modifying their secondary.

(48) He may have heard us, but if so he did not show it SVO&3(3^c3)SV^nO_2 (*AS*: 72)

Here we have a *part of speech*, namely *so*, preceded by a connective (a part of speech usually without a rank of its own), namely *if*.

(49) He was usually well-dressed SV3P($3Y^b$) (*AS*: 32)

For Y the convention used by Jespersen is the following: '[w]hen nothing is said expressly of the rank of Y, it is a primary; if it is secondary the symbol 2(Y) is used except when Y is accompanied by a 3, which implies that Y is itself 2' (1937: 147). Here ($3Y^b$) is a quasi-nexus saturated by its subject *He* S.

(50) All the world over 1(21)p (*AS*: 33)
There is nothing very serious the matter 3/sVS(132)P (*AS*: 33)

Above we have examples of junctions where the usual order is altered, which may be called *irregular junctions.*

(51) He had a shilling a week pocket money SVO(13[2–1]) (*AS*: 36)

In brackets we have the very suspicious jump in numeration 1 3. In my view *a week* is not a tertiary as in *He had a shilling a week* SVO3, but is a secondary which modifies *a shilling*, and *pocket money* is an apposition of the whole. Hence the formula SVO(1 2) [2–1].

(52) Their first talks alone together on the balcony S^2*2X[23p1]* (*AS*: 36)

The brackets contain a *quasi-nexus* in which p1 is a second tertiary, if supposed to modify *alone*, and its subject is the form *they* represented by S, because S^2 represents *their*, its genitive.

(53) They came back with Jones as spokesman SV3p1(S_2pP) (*AS*: 43)

Here pP is not tertiary as in (39a) but secondary of the nexus S_2pP, so it is better to write (S_2P(p1)).

(54) He made believe to weep $SVO^r(S^o{}_2IO_2(S^oI))$ (*AS*: 44)

In (54) the quasi-nexus *believe to weep* has a special character because it is saturated by the understood subject of *believe*.

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

(55) (a) You are the one to speak to the Minister SVP(12(Ip1)) (*AS*: 49)
(b) You are to speak to the Minister SVP(2=Ip1) (*AS*: 49)

In (55a) (Ip1) is a quasi-nexus which is saturated by *the one*, the primary of the junction *the one to speak to the Minister* 12(Ip1), while in (55b) (Ip1) is a quasi-nexus which is saturated by *you* and the whole *you to speak to the Minister* is the nexus SIp1. Why is the first expression a junction and the second a nexus? Because *the one* is indefinite and the function of *to speak to the Minister* is to define it, while *you* is quite definite in itself and what follows is an additional idea, namely what is predicated of *you*. The p1 *to the Minister* is a prepositional indirect object and is better rendered by pO^{i}.

(56) This is a delightful room in which to work SVP(212($p1^{c}pI$) (*AS*: 52)

In (56) ($p1^{c}pI$) is an infinitival relative clause, therefore a nexus, better a quasi-nexus which has the special character of being saturated by a latent subject which is the 'generic person', indicated by Jespersen by the symbol S^{∞} (Jespersen, 1937: 128). A more meaningful formula would be ($3(p1^{c})S^{\infty}$ ^{o}pI). But why pI and not simply I?

(57) This is too good to be true SVP ($32p1(IP_2)$) (*AS*: 54)

Jespersen treats *to be true* as a tertiary qualifying *good*. But does *to be true* modify *good*? In my view this is another case of a split subject, namely $½S(S_1)VP(32)½S(IP_1)$ where *to be true* is a quasi-nexus having *this* S_1 as its subject and *too good* (32) is a quasi-nexus having *this to be true* $½S(S_1)½S(IP_1)$ as its subject.

(58) He will give an account of why he did it SVOp1(3^{c}? SVO) (*AS*: 64)

Is *of why he did it* p1(3^{c}?SVO), a prep-nexus having as its regimen an interrogative clause, tertiary? In my view this prep-nexus is the object of the nexus-substantive *an account*. The formula becomes $SVO(XpO_1(3^{c}?SVO))$.

(59) I know the family secrets, young as I am S*VO(2–1)[*$P3^{c}SV$] (*AS*: 67)

The rank of $3^{c}SV$ is left undecided. The apposition is a quasi-nexus saturated by *I* S*. Better write S*VO(2–1)[P3($3^{c}SV$)]*.

(60) We burst out laughing when we saw him SV3Y3($3^{c}SVO$) (*AS*: 69)

The participle *laughing* Y is a secondary like *burst out* V3 and also a predicative P of the subject, and the tertiary 3($3^{c}SVO$) modifies *burst out laughing* as a whole. The two secondaries not being independent, I would use the square brackets [ ] (see section 3.3 for this use of [ ]) to express the relation between the two connected verbal forms, and write the formula $SV(V_1{}^{*}3[P(Y)]^{*})3(3^{c}SVO)$.

(61) Things move so fast that we cannot follow SV433($3^{c}S_2V^{n}$) (*AS*: 69)

The nexus in brackets modifies *fast* and consequently is 4: SV434($3^{c}S_2V^{n}$).

(62) The older he grows, the more he drinks 3($3^{c}PSV$)O(31)SV (*AS*: 71)

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

In (62) the articles are represented because they correlate two clauses. The symbol 3^{c} is used for the article in the ordinate, while in the superordinate the article is assigned its appropriate rank, which is 2 and not 3. But what about the rank of the ordinate? I would use the square brackets [ ] and write [3^{c}PSV§]O(21)SV.

Summarizing the results of this survey, we can say that what is enclosed in brackets may be:

– strings of variously connected or unconnected basic forms (parts of speech)
– junctions/nexuses in all their various forms illustrated above

Between the intermediate level and the sentence level, one might envisage another level of representation, that occupied by all the discontinuous constituents which form a hierarchical scale and can be extracted from the sentence in a way semantically compatible with the sentence. These constituents may be called *macro-constituents* and a sentence containing them may be considered the result of a merging of its macroconstituents. Let me illustrate this macroconstituent level of representation with some examples:

(63) He arrived home late SV3(33)

In (63) there are two macroconstituents: *He arrived home* SV3 and *He arrived late* SV3.

(64) How kind of you to come! P(3!2)p1(S_2*)S(I*) (*AS*: 41)

The macroconstituents are: *How kind of you you* P(3!2)3(p1)S_2*, *How kind you to come* P(3!2)S(S_2I).

(65) Her temperature had dropped to normal S(1^{2}1)VpP^{r} (*AS*: 30)

Macroconstituents: *normal temperature* P^{r}(21), *Her temperature had dropped to normal temperature* S(1^{2}1)Vp 3(pP^{r}(21)).

(66) We feed children who(m) we think are hungry SVO(12(O^{c}/$S^{c}{}_{2}$SVV_2P)) (*AS*: 68)

Here we have: *We feed children* SV1, *we think children are hungry* S$V1V_2$P. The following case is different:

(67) I argued against everything he believed had happened
SWO(S_2*S_3VO_2(V*))[7] (*AS*: 68)

where the intermediate-level constituent (S_2*S_3VO_2(V*)) is the junction 1(S_2*)2(S_3VO_2(V*)) in which the secondary is a nexus having a quasi-nexus as object which is saturated by the primary of the junction. Only in certain cases may a sentence be interpreted as an interweaving of macroconstituents.

[7] On 'W', see p. 139 above.

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

3 A modernized version of the algorithm

Modern linguists may not feel completely at ease with some imperfections of Jespersen's formalism. Here I will deal with some possible refinements of it. The new formulas will be marked by the abbreviation MV (modernized version), to distinguish them from their original *AS* versions.

3.1 Representation of the article

The *AS* version would allow the representation of the article by means of the symbol of rank 2 but there is no symbol in (25c) of section 2.2 which could be used properly to describe it as a form. This list should be extended to include the symbol 'det' for determinative, thus giving the algorithm the possibility of marking the article or other determinative when required by a more detailed analysis. However, a systematic representation of the article would make the formulas more cumbersome without any added advantage. An optional representation, which would require removing (25b) in section 2.2 and replacing it by (25)(b) <article>: : = <optional>, might be desirable in a limited number of cases.

Niels Haislund, Jespersen's secretary from 1928 to 1943, the year of his death, and the collaborator who assisted him, together with Paul Christophersen, in writing *MEG* (VI) and, after his death, completed *MEG* (VII), wrote in the latter:

> It is regrettable that Jespersen did not consider the articles in his analyses in Analytic Syntax, for then he would have been forced to decide whether the articles should be considered as belonging to the primary of a junction or to the junction as a whole. (Jespersen, 1949: 471)

Consider such junctions as

(68) a certainly not very cleverly worded remark 32(5432)1 (*AS*: 111 note 1)
a federal government inquiry 2(21)1

These might be rendered, by a modern linguist who wants to include the article, by one of the two alternative formulas

(69) a certainly not very cleverly worded remark 2(det)32(5432)1/2(det)1(32(5432)1) (MV)
a federal government inquiry 2(det) 2(21)1/2(det)1(2(21)1) (MV)

3.2 Use of round brackets and braces

The round brackets usually follow a symbol of function, and what is inside them usually describes the form which fills the function. But there are cases in *AS* where this practice is not observed:

(70) John told Robert's son that he must help him $SVO^{i}(1^{2}1)O(3^{c}SVO(O^{i}))$ (*AS*: 26)
He was not himself that night $SV^{n}P(S)3(21)$ (*AS*: 28)

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

In (70) round brackets are used in the items $O(O^{i})$ and P(S) to express the fact that two different items are coreferential. Their use may be easily obviated simply by introducing a *coreference index* j, k, etc. as a subscript:

(71) John told Robert's son that he must help him $SVO^{i}_{j}(1^{2}1)O(3^{c}SVO_{j})$) (MV)
He was not himself that night $S_jV^n\ P_j3(21)/\ S_jV^n\ P_j\ (1)3(21)$ (MV)

In the example

(72) They hate one another $SVO(S_x)$ (*AS*: 6)

the subscript x may be used as a *reciprocity index* in a way similar to a coreference index

(73) They hate one another S_xVO_x (11) (MV)

In

(74) The butler was offered a reward $S(O^{ia})V^{b}O$ (*AS*: 26)

the symbolization $S(O^{ia})$ for *the butler* means that what is the indirect object in the active may become subject in the passive transform, i.e. *the butler* shares two functions. Whenever we want to describe an item sharing more than one function in the same sentence we could list all its functions, starting from the most significant, inside the braces { }, as Jespersen does for the second person of the imperative in *Come now!* $\{SV\}3!^{r}$ (*AS*: 8), a formula which I would change into $\{VS\}3!^{r}$ (MV). After the braces we could optionally describe the form in round brackets, whenever it seems necessary:

(75) The butler was offered a reward $\{SO^{ia}\}V^{b}O/\{SO^{ia}\}(2(det)1)V^{b}O$ (MV)

Similarly in

(76) You have nothing to fear $SVO(1(O_2*)2(I*))$ (*AS*: 51)
You have nothing to fear $SVO(\{1O_1*\}2(I*))$ (MV)

where the infinitive *to fear* I* is *retroactive*, i.e. it is interpreted also as having an object before its occurrence in the sentence.

In the case of split subject/object as in

(77) She was seen to run $½SV^{b}½S(I)$ (*AS*: 46)
This was considered a lie $½SV^{b}½S(P)$ (*AS*: 46)

I would preserve the round brackets, because (I) and (P) are parts of the forms *She to run* S_1I and *This a lie* S_1P and both these forms are present in the sentences. But I would change the formulas into

(78) She was seen to run $½S(S_1*)V^{b}½S(I*)$ (MV)
This was considered a lie $½S(S_1*)V^{b}½S(P*)$ (MV)

to differentiate S_1s, the grammatical subjects of V^{b}s and the subjects of the quasi-nexuses I and P from the notional subjects Ss of V^{b}s *She to run, This a lie*. Nevertheless there would be no harm in writing

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

(79) She was seen to run {½SS$_1$*}V^{b}½S(I*) (MV)
This was considered a lie {½SS$_1$*}V^b{½SP*} (MV)

In

(80) An utter fool 2(3) 1 (*AS*: 7)[8]

the symbol '2(3)' means that *utter* is a secondary which is also a tertiary to the adjective *foolish* contained in *fool*, so I suggest changing the formula into

(81) An utter fool {23}1 (MV)

The following examples

(82) This fruit is good to eat S(O*)(21)VP(2pI*) (*AS*: 52)
This is a hard nut to crack SVP(21)(O*pI*) (*AS*: 52)

which present two consecutive pairs of round brackets, are unintelligible to me, and I am inclined to think that they, together with some other four or five in the same section, originated from printing errors. I would revise them as follows:

(83) This fruit is good to eat {SO*}(21)VP(23(I*)) (MV)
This is a hard nut to crack SVP(O*(2**1)3**(I*)) (MV)

3.3 Use of square brackets

Jespersen used the brackets [ ] for appositions, i.e. for items which have a rank nearly but not fully equipollent to the item which the appositions refer to, for if the rank were equipollent the square brackets would not be necessary. They express the kind of loose hierarchical relation that holds between an item and its apposition, and in this way they are a sort of *operator* which produces a degrading of the rank of what the apposition refers to, and assigns this degraded rank to it. We could give this more general meaning of *rank degrader* to the sign [ ], and use it whenever it might appear desirable to introduce an intermediate rank in the interval below:

0 1 2 3
|— (sentence, no rank) —|— (S,O,pO etc . . .) —|— (V or I,G,Y or P) —| (tertiaries) — . . .

Let us next consider the sentence

(84) John gave Mary a kiss in a hurry SVO^{i}O3(p1)

Jespersen identifies the ranks of *Mary* and *a kiss* as primaries, and calls them both primaries, because both select/are selected by the secondary *gave*, but they are not equipollent, for we can have *John gave a kiss in a hurry* but we cannot have **John gave Mary in a hurry*, and for this reason there must be a hierarchical relation between *Mary* and *a kiss*, because the occurrence of Mary is conditioned by *a kiss*. We might call *Mary* a subprimary, assign it the conventional decimal value 1.1 and

[8] Different from *the mentally deficient* 2(3)1 (*AS*: 114) and *In perfect good temper* 3(2)21 (*AS*: 9), where the brackets describe forms.

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

represent it in the formula by [O^i]. A subprimary, in so far as it is primary, is connected with the verb, which is secondary, and in so far as it is subprimary 1.1, is connected with the primary O but not with the primary S, from which it is separated by the verb. The formula becomes

(85) John gave Mary a kiss in a hurry SV[O^i]O3(p1) (MV)

In the sentence

(86) He married young and died poor SVP&V_2P_2 (*AS*: 36)

we could use the square brackets to specify *young* and *poor* as subsecondary 2.1 and write

(87) He married young and died poor SV[P]&V_1[P_1] (MV)

where both the predicatives as 2 modify the subject and as 2.1 modify the verb. Here, the rank 3 being available, an alternative formula would be

(88) He married young and died poor SV{P3}&V_1{$P_1$3} (MV)

In the sentence

(89) I saw her, however SVO[§3]

however refers to the whole sentence, hence the use of the sign §, and the square brackets assign the subzero rank 0.1 to the adverb 3 *however*. In the sentence

(90) The letter, which Kim had written, was posted yesterday S(1[O^cS_1V])V^b3 (MV)

the square brackets assign the relative clause, which is nondefining and as such inessential as an apposition, the rank 1.1 of subprimary of the primary *The letter*, and in this way we can differentiate defining relative clauses, which are secondary, from nondefining ones. (See (30) in section 2.2 for an example of restrictive relative clause.) In the sentence

(91) John put the car in the garage SVOpO_1

the two objects, the direct and the prepositional one, go together, for neither of them may stand by itself. I would claim that pO_1 is not tertiary because it is also conditioned by O, therefore it is in hierarchical relation with O which is primary and consequently it is a subprimary. The sequence OpO_1 obviously has a structure, but it would be otiose to analyse it by using the sign [ ]. (One could split hairs and say that in a sense both the objects are subprimaries because they are conditioned by one another and the hierarchical relation between them could be expressed by assigning the first the conventional value 0.9 and the second the conventional value 1.1: both being decimal, neither one is independent.)

I see a difference between *put* in (91) and *find* in

(92) John found the car in the garage SVO3(p1)

In (92) *the car* and *in the garage* can be dissociated, *John found the car* being grammatical, and *in the garage* modifies the secondary *found* as a tertiary/adverbial.

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

In (91) *in the garage* cannot be assigned the tertiary rank. I see two possibilities: either *the car in the garage* is a whole which is in syntactic relation with *put*, or *put in the garage* is a whole (something equivalent to *garaged*) which is in syntactic relation with *the car*. In the first case what is said above holds good and *in the garage*, being linked to the primary/object *the car*, becomes a subprimary, in the second *in the garage* belongs to the verbal position, i.e. to the secondary, in that it completes the sense of the verb *put* which, without it, has no meaning and cannot have an object. In neither of these cases is *in the garage* adverbial, if adverbial means tertiary, something which can modify, independently of other elements, a verb. In my view, syntax is concerned with free combinations of words or word groups. If there is something obligatory in a sentence, first syntax must take into account the block formed by the words which cannot be separated and decide what its relation is with the rest of the sentence. In (92) this block is *found the car* and its relation with *John* is 12 (this is the only possible dependence between a subject and a block of words related to it), while its relation with *in the garage* is 23. In (91) the block is *put the car in the garage* and its relation with *John* is still 12. Second, syntax must analyse the blocks and decide what is the governing/obliging part and what is the governed/obliged part of the whole. There is no problem for the whole *found the car* VO and our complete analysis becomes

John found the car in the garage 1=S 2=VO 3=p1/SVO3(p1)

For the whole *put the car in the garage* there are the two choices discussed above. They are expressed by the formulas

John put the car in the garage 1=S 2=VOpO_1/SVOpO_1
John put the car in the garage 1=S 2=½VO½V(p1)/S½VO½V(p1)

There are cases in which the function of completing the meaning of the verb is fulfilled by an adjective (*The government set the prisoners free*) and by a noun (*She called her baby George*). In none of these or similar examples does the term adverbial seem to me to be adequate or appropriate, either functionally or formally.

The two sentences

(93) They proved him guilty SVO(S_1P)
They pushed the door open SV*O_i[$S^o{}_iP^r$]*/SV{OS_1*}{P^r*3} (MV)

between which there is a significant difference, namely that *proved him guilty* does not involve *proving him*, but *pushed the door open* does involve *pushing the door*, are well differentiated by their formulas. The square brackets separate *open* from the object *the door* of *pushed*, and at the same time relate it to *the door* as a predicative of result and by means of the asterisks put it in hierarchical relation also with V by degrading its rank to a subsecondary. The application of the sign [ ] has the effect of shifting the rank of P(2) to the conventional value 2.1, which allows P^r to have a double hierarchical relation, one with *the door* (rank 1) as predicative and another with *pushed* (rank 2) as a modifier. The alternative formula expresses the same facts in another way.

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

3.4 Indefinite rank

When an item has a rank which is different from that implied by the theory, this rank must always be specified. In particular, prepositional groups and infinitives, all of which may variously be primaries, secondaries and tertiaries, must always have their rank explicitly expressed. In the sentences:

(94) They laughed at Jim SVpO
They went to Paris SVpO
The paper remained with Jim SV3(p1)
John was found in the garden $SV^{b}3(p1)$

the fact that *at Jim* is primary is expressed by using O instead of 1, while the fact that *with Jim* is tertiary is rendered by 3(p1). Why is *at Jim* primary and *with Jim* tertiary? Because *at* belongs to the verb *laugh*, is selected by the verb and contributes to its meaning, while *remain* has a meaning independent of the preposition that follows it and may be followed by many different prepositions (*The paper remained on the table/under the table/behind the bush, etc*.).

3.5 To-infinitive and bare-infinitive

In the sentences

(95) We saw him run $SVO(S_2I)$ (*AS*: 43)
I made him want to go to France $SVO^{r}(S_2IO_2(Ip1))$ (*AS*: 44)
The motor car has come to stay SVpI (*AS*: 53)

the symbol I represents the bare-infinitives *run, want*, the to-infinitive *to go*, and the base form *stay* of the to-infinitive *to stay* (which is represented by pI because it is tertiary). I would suggest using I systematically for to-infinitives, I_b for bare-infinitives (also for didactic purposes) and representing a tertiary to-infinitive by 3(I). The formulas become

(96) We saw him run $SVO(S_1I_b)$ (MV)
I made him want to go to France $SVO^{r}(S_1I_bO_1(IpO_2))$ (MV)
The motor car has come to stay SV3(I) (MV)

3.6 Summary of the extensions

The MV version of the algorithm requires the following changes in section 2.2:

(25) (b) <article>: : = <optional>
(c) <form>: : = det|V|v|I|G|X|Y|Z|p|pp|&|3^{c}|3^{cc}|3/s|s|o|o^{i}|+

Forms:

det: determinative
V: finite verb, etc . . .

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

Brackets:

( ) What is enclosed in round brackets is always a form which realizes the function (or functions) of the item that precedes it or, complementarily, an item followed by the opening sign '(' is always a function filled by the form which begins after it and finishes with its corresponding closing sign ')'.

[ ] Square brackets are used for extraposition or apposition and as *rank degrader*. Their effect is to degrade the rank of their referent, assign this degraded rank to what is contained in the brackets and establish, in this way, a hierarchical relation between them. If the referent is a primary, the degraded rank is a subprimary (conventionally indicated by the decimal value 1.1 or 0.9); if the referent is a secondary, the degraded rank is a subsecondary (decimal value 2.1); if the referent is a sentence (no rank, conventional rank 0), the degraded rank is a subzero rank (decimal value 0.1).

{ } The braces are used when an item has two or more functions. They contain the list of all the functions shared by the item, headed by the most significant, and may be followed by a pair of round brackets which describes the form of the item.

2^{m} : secondary of an unclassifiable junction

. . .

S_j, O_j . . . : coreferential functions

S_x, O_x . . . : reciprocal functions

I_b : bare-infinitive

4 Conclusion

In *Syntactic Structures* (Chomsky, 1957), the publication which is generally recognized as a turning point in twentieth-century linguistics, Chomsky wrote

> In each case our sole concern has been to decrease the complexity of the grammar . . . Making use of phrase structure and transformations, we are trying to construct a grammar of English that will be simpler than any proposed alternative; and we are giving no thought to the question of how one might actually arrive at this grammar in some mechanical way from an English corpus, no matter how extensive . . . The intuitive correspondences and explanations of apparent irregularities seem to me to offer important evidence for the correctness of the approach. (Chomsky, 1957: 83–4)

> It is undeniable that 'intuition about linguistic form' is very useful to the investigator of linguistic form (i.e. grammar). It is quite clear that the major goal of grammatical theory is to replace this obscure reliance on intuition by some rigorous and objective approach. There is, however, little evidence that 'intuition about meaning' is at all useful in the actual investigation of linguistic form. (Chomsky, 1957: 94)

It is strange that with these beginnings it fell to the lot of the Chomskyan school to design a formal apparatus which has a degree of abstraction and complexity which stands out in sharp contrast with the characteristics of simplicity, intuition, manageability, and versatility that may be found in Jespersen's formalism. Chomsky's apparatus includes PS rules, tree diagrams, labelled bracketings, two-dimensional graphs with straight lines, triangles, circles, arrows, etc., and a variety of codified notations for subcategorization frames, categorial features, transformations, etc. Chomskyan grammars deal with discontinuities (relations between distant words) by

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

means of transformations which imply processes which usually are not formalized, and what is represented is only the labelled bracketing of their initial and final stage, while JG translates relations into hierarchical scales between words, not necessarily contiguous, and uses asterisks for clarifications in case of ambiguity.

Many modern systems of different persuasions borrowed from Chomsky the phrase structure description and the tree diagram. They recognize the existence of at least six intermediate constituents, namely VP, NP, AdjP, AdvP, PP, DetP, most of them having a fixed and rigid structure, and all of them formed by contiguous words. JG has only two intermediate constituents, junction and nexus, having a flexible and variable structure, not necessarily formed by contiguous words.

The hierarchical construction inherent in the tree diagram differs substantially from that of Jespersen. In the tree diagram the hierarchical relations are generated by a top–down splitting process which is regulated by the recognition of the constituents theorized about by the grammar applied to the analysis. Consequently they have an abstract (non-intuitive) and derivative character, in so far as they are determined by a complex theory which is *magna pars* of a grammar, the theory of constituents including also the parts of speech. By contrast, Jespersen's ranks are intuitive, primitive, natural. They are based on the relations existing in the series adverb–adjective–noun (e.g. *furiously barking dogs*) made up by three fundamental forms of a language. They are extended just as naturally to the nexus (e.g. *dogs barked furiously*). Syntactic analysis consists of discovering these ranks inside a sentence, which is taken in all its entirety, without any preliminary and possibly arbitrary splitting, starting down–top from the words and introducing a junction or a nexus only when a rank is served not by one word, but by a group of words. The modernized version of Jespersen's algorithm allows the insertion of intermediate ranks between and beyond the three classical ranks and, as it were, realizes the continuum of ranks in the sentence domain. The formalism is so flexible that different degrees of approximation may be used in the formulas.

All the examples dealt with in this paper cover many different aspects of syntax. The application of Jespersen's algorithm to their exploration gives clear evidence of its power of analysis, its pedagogical value and its deep insight into grammatical constructions. The pedagogical value is corroborated by innumerable disclosures of recondite syntactic features that catch your eye only when you follow the sentence word for word in order to decide about their rank, form, and function. In many cases linguistic structures turn out to be better explored theoretically by Jespersen's grammar than by more modern systems.

Author's address:
Corso Massimo D'Azeglio 10
10125 Torino
Italy
s.digiorgi@libero.it

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press

References

Backus, J. W. (1959). The syntax and semantics of the proposed international algebraic language of the Zürich ACM-GAMM conference. *ICIP*: Paris, June 1959.
Bodelsen, C. A. (1964). *Essays and papers*. Copenhagen.
Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
Di Giorgi, S. (1997). Jespersen's v. Huddleston's syntactic analysis. RASK **7**: 3–47.
Huddleston, R. (1988). *English grammar, an outline*. Cambridge: Cambridge University Press.
Jespersen, O. (1909–49). *A modern English grammar on historical principles (MEG*), vol. 7 (1949). London: Allen & Unwin.
Jespersen, O. (1936). *Specimens of syntactic formulas*. Copenhagen: Levin & Munksgaard.
Jespersen, O. (1937). *Analytic syntax* (*AS*). New York: Holt, Rinehart & Winston.
Jespersen, O. (1995). *A linguist's life*. Odense: Odense University Press.
Juul, A. & H. F. Nielsen (1989). *Otto Jespersen: facets of his life and work*. Amsterdam and Philadelphia: John Benjamins.
McCawley, J. D. (1970). Review of *Analytic Syntax*. *Language* **46**: 442–9.
McCawley, J. D. (1991). Review of Juul & Nielsen, 1989. *Language* **67**: 117–20.
Quirk, R. (1994). Exploring the English genitive. In *A linguistic miscellany. Proceedings of the Otto Jespersen Symposium, April 29–30, 1993*. Department of English, University of Copenhagen. 7–22.

https://doi.org/10.1017/S1360674301000168 Published online by Cambridge University Press