# CURS 6 Fișiere text

Un *fișier text* este o secvență de caractere organizată pe linii și stocată în memoria externă (SSD, HDD, DVD, CD etc.). Practic, fișierele text reprezintă o modalitate foarte simplă prin care se poate asigura *persistența datelor*.

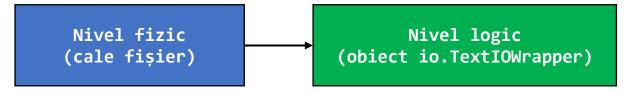
Liniile dintr-un fişier text sunt demarcate folosind caracterele **CR** (*carriage return* sau '\r' sau chr(13)) şi **LF** (*line feed* sau '\n' sau chr(10)), astfel:

- în sistemele de operare *Windows* și *MS-DOS* se utilizează combinația **CR+LF**;
- în sistemele de operare de tip *Unix/Linux* și începând cu versiunea *Mac OS X* se utilizează doar caracterul **LF**;
- în sistemele de operare până la versiunea *Mac OS X* se utilizează doar caracterul **CR**.

Indiferent de modalitatea utilizată pentru demarcarea liniilor într-un anumit sistem de operare, în momentul în care se citesc linii dintr-un fișier text, caracterele utilizate pentru demarcare vor fi transformate automat într-un singur caracter LF ('\n')!

## Deschiderea unui fisier text

Operația de deschidere a unui fișier (text sau binar!) presupune asocierea fișierului, considerat la nivel fizic (i.e., identificat prin calea sa), cu o variabilă/un obiect dintr-un program, ceea ce va permite manipularea sa la nivel logic (i.e., prin intermediul unor funcții sau metode dedicate):



Practic, după deschiderea unui fișier, obiectul asociat fișierului va conține mai multe informații despre starea fișierul respectiv (dimensiunea în octeți, modalitatea de codificare utilizată, poziția curentă a pointerului de fișier etc.), iar programatorul va putea acționa asupra fișierului într-un mod transparent, folosind metodele puse la dispoziție de obiectul respectiv.

Deschiderea unui fișier se realizează folosind următoarea funcție (reamintim faptul că parantezele drepte indică faptul că parametrul este opțional):

Primul parametru al funcției reprezintă calea fișierului, sub forma unui șir de caractere, iar în cazul în care fișierul se află în directorul curent se poate indica doar numele său. Modul de deschidere este utilizat pentru a preciza cum va fi prelucrat fișierul text și se indică prin următoarele litere:

- "r" pentru citire din fișier (read)
  - fișierul va putea fi utilizat doar pentru a citi date din el;
  - este modul implicit de deschidere a unui fișier text, i.e. se va folosi dacă se omite parametrul "mod de deschidere" în apelul funcției open;
  - dacă fișierul indicat prin calea respectivă nu există, atunci se va genera eroarea FileNotFoundError;
- "w" pentru scriere în fișier (write)
  - fisierul va putea fi utilizat doar pentru a scrie date în el;
  - dacă fișierul indicat prin calea respectivă nu există, atunci se va crea unul nou, altfel fișierul existent va fi șters automat și va fi înlocuit cu unul nou;
- "x" pentru scriere în fișier creat în mod exclusiv (exclusive write)
  - fișierul va putea fi utilizat doar pentru a scrie date în el;
  - dacă fișierul indicat prin calea respectivă nu există, atunci se va crea unul nou, altfel se va genera eroarea FileExistsError;
- "a" pentru adăugare în fișier (append)
  - fișierul va putea fi utilizat doar pentru a scrie informații în el;
  - dacă fișierul există, atunci noile datele se vor scrie imediat după ultimul său caracter, altfel se va crea un fișier nou și datele se vor scrie de la începutul său.

Pentru a trata excepțiile care pot să apară în momentul deschiderii unui fișier, se va folosi o instrucțiune de tipul try...except, așa cum se poate observa în exemplele următoare:

```
try:
    f = open("exemplu.txt")
    print("Fișier deschis cu succes!")
except FileNotFoundError:
    print("Fișier inexistent!")
```

```
try:
    f = open("C:\Test Python\exemplu.txt", "x")
except FileNotFoundError:
    print("Calea fișierului este greșită!")
except FileExistsError:
    print("Fișierul deja există!!")
```

Din cel de-al doilea exemplu se poate observa faptul că eroarea FileNotFoundError va fi generată și în cazul modurilor de deschidere pentru scriere (i.e., modurile "w", "x" și "a") în cazul în care calea fișierului este greșită!

# Citirea datelor dintr-un fișier text

În limbajul Python, datele citite dintr-un fișier text vor fi furnizate întotdeauna sub forma unor șiruri de caractere!

Pentru citirea datelor dintr-un fișier text sunt definite următoarele metode:

• read(): furnizează tot conținutul fișierului text într-un singur șir de caractere

#### Exemplu:

Program	Fișier de intrare	Ecran
<pre>f = open("exemplu.txt") s = f.read()</pre>	Ana are multe pere,	Ana are multe pere,
		mere rosii, si mai multe prune!!!

Observați faptul că șirul s conține și caracterele LF ('\n') folosite pentru delimitarea liniilor!

• **readline()**: furnizează conținutul liniei curente din fișierul text într-un șir de caractere sau un șir vid când se ajunge la sfârșitul fișierului

## Exemplu:

F	1	ī
Program	Fișier de intrare	Ecran
<pre>f = open("exemplu.txt") s = f.readline()</pre>	Ana are multe pere,	Ana are multe pere,
	mere rosii, si mai multe prune!!!	mere rosii, si mai multe prune!!!
<pre>s = f.readline() f.close()</pre>		

Observați faptul că, de fiecare dată, șirul s conține și caracterul LF ('\n') folosit pentru delimitarea liniilor, mai puțin în cazul ultimei linii!

O modalitate echivalentă de parcurgere a unui fișier text linie cu linie constă în iterarea directă a obiectului f asociat:

#### Exemplu:

Program	Fișier de intrare	Ecran
for linie in f:	, ,	Ana are multe pere,
<pre>print(linie, end="") f.close()</pre>	mere rosii, si mai multe prune!!!	mere rosii, si mai multe prune!!!

 readlines(): furnizează toate liniile din fișierul text într-o listă de șiruri de caractere

#### Exemplu:

Program	Fișier de intrare	Ecran
<pre>f = open("exemplu.txt") s = f.readlines()</pre>	Ana are multe pere,	<pre>['Ana are multe pere,\n', '\n', 'mere rosii,\n', 'si mai</pre>
<pre>print(s) f.close()</pre>	mere rosii, si mai multe prune!!!	rosii,\n', 'si mai multe prune!!!']

Observați faptul că fiecare șir de caractere din listă, corespunzător unei linii din fișier, conține și caracterul LF ('\n') folosit pentru delimitarea liniilor, mai puțin în cazul ultimei linii!

Cele 3 modalități de citire a datelor dintr-un fișier text sunt echivalente, dar, în cazul în care dimensiunea fișierului este mare, se preferă citirea linie cu linie a conținutului său, deoarece necesită mai puțină memorie.

Deoarece toate metodele folosite pentru citirea datelor dintr-un fișier text furnizează șiruri de caractere, dacă vrem să citim datele respective sub forma unor numere trebuie să utilizăm funcții pentru manipularea șirurilor de caractere.

**Exemplu:** Fişierul text *numere.txt* conține, pe mai multe linii, numere întregi separate între ele prin spații. Scrieți un program care afișează pe ecran suma numerelor din fisier.

#### **Soluția 1** (folosind metoda read):

Construim o listă numere care să conțină numerele din fișier, împărțind șirul corespunzător întregului conținut al fișierului în subșiruri cu metoda split și transformând fiecare subșir într-un număr cu metoda str:

```
f = open("numere.txt")
numere = [int(x) for x in f.read().split()]
print("Suma numerelor din fisier:", sum(numere))
f.close()
```

#### **Solutia 2** (folosind metoda readline):

Construim tot o listă care să conțină toate numerele din fișier, dar împărțind în subșiruri doar șirul corespunzător liniei curente:

```
f = open("numere.txt")
numere = []
linie = f.readline()
while linie != "":
    numere.extend([int(x) for x in linie.split()])
    linie = f.readline()
print("Suma numerelor din fisier:", sum(numere))
f.close()
```

Evident, o soluție mai eficientă din punct de vedere al memoriei utilizate constă în calculul sumei numerelor de pe fiecare linie și adunarea sa la suma tuturor numerelor din fișier:

```
f = open("numere.txt")
total = 0
linie = f.readline()
while linie != "":
    numere = [int(x) for x in linie.split()]
    total = total + sum(numere)
    linie = f.readline()
print("Suma numerelor din fisier:", total)
f.close()
```

### **Soluția 3** (folosind metoda readlines):

Vom proceda într-un mod asemănător cu soluția 2, dar parcurgând lista cu liniile fișierului furnizată de metoda readlines:

```
f = open("numere.txt")
total = 0
for linie in f.readlines():
    numere = [int(x) for x in linie.split()]
    total = total + sum(numere)
print("Suma numerelor din fisier:", total)
f.close()
```

Atenție, în oricare dintre soluțiile prezentate, metoda split trebuie apelată fără parametri, pentru a împărți șirul respectiv în subșiruri considerând toate spațiile albe (care includ caracterele '\n' și '\r')! Altfel, dacă am apela metoda split doar cu parametrul " "(un spațiu) ar fi generată eroarea ValueError în momentul în care am încerca să transformăm în număr ultimul subșir de pe o linie (mai puțin ultima din fisier), deoarece acesta ar contine si caracterul '\n'!

# Scrierea datelor într-un fișier text

În limbajul Python, într-un fișier text se pot scrie doar șiruri de caractere. Pentru scrierea datelor într-un fișier text sunt definite următoarele metode:

• write(sir): scrie șirul respectiv în fișier, fără a adăuga automat o linie nouă

#### Exemplu:

Program	Fișier de ieșire
<pre>f = open("exemplu.txt", "w") cuvinte = ["Ana", "are", "mere!"] for cuv in cuvinte:     f.write(cuv) f.close()</pre>	Anaaremere!

Dacă dorim să scriem fiecare cuvânt pe o linie nouă, atunci trebuie să modificăm f.write(cuv)în f.write(cuv + "\n").

• writelines(colecție de șiruri): scrie toate șirurile din colecție în fișier, fără a adăuga automat linii noi între ele

## Exemplu:

```
Program

f = open("exemplu.txt", "w")
cuvinte = ["Ana", "are", "mere!"]
f.writelines(cuvinte)
f.close()
Anaaremere!
```

Dacă dorim să scriem fiecare cuvânt pe o linie nouă, atunci fie trebuie să adăugăm câte un caracter "\n" la sfârșitul fiecărui cuvânt din listă (de exemplu, prin instrucțiunea cuvinte = [cuv + "\n" for cuv in cuvinte]), fie să modificăm f.writelines(cuvinte) în f.writelines("\n".join(cuvinte)), deci să concatenăm toate șirurile din listă într-unul singur, folosind caracterul "\n" ca separator (evident, în acest caz putem utiliza și metoda write).

# Închiderea unui fișier text

Indiferent de modalitatea în care a fost deschis un fișier și, implicit, accesat conținutul său, acesta trebuie închis după terminarea prelucrării sale, folosind metoda close(). Închiderea unui fișier constă în golirea eventualei zone tampon (buffer) alocate fișierului și ștergerea din memorie a obiectului asociat cu el. Dacă un fișier deschis într-unul dintre modurile de scriere nu este închis, atunci există riscul ca ultimele informații scrise în fișier să nu fie salvate! În mod implicit, un fișier este închis dacă referința obiectului asociat este atribuită unui alt fișier (i.e., se utilizează, din nou, funcția open).

Putem elimina necesitatea închiderii explicite a unui fișier putem folosi o instrucțiune with...as, astfel:

```
with open("exemplu.txt") as f:
    s = f.readlines()
    print(f.closed) #False

print(f.closed) #True
print(s)
```

În acest caz, fișierul va fi închis automat imediat după ce se va termina prelucrarea sa, așa cum se poate observa din exemplul de mai sus (data membră closed permite testarea stării curente a unui fișier, respectiv dacă acesta a fost închis sau nu). Mai mult, închiderea fișierului respectiv se va efectua corect chiar și în cazul apariției unei erori în timpul prelucrării sale. Atentie, chiar dacă se utilizează o instructiune

with...as, erorile care pot să apară în momentul deschiderii unui fișier trebuie să fie tratate explicit!

În încheiere, vom prezenta un exemplu în care vom utiliza toate cele 3 moduri de deschidere ale unui fișier text.

Mai întâi, vom implementa un program care să genereze un fișier text care va conține pe prima linie o sumă formată din n numere naturale aleatorii (de exemplu, 12 + 300 + 9 + 1234):

```
import random

numef = input("Numele fisierului: ")
n = int(input("Numarul de valori aleatorii: "))

with open(numef, "w") as f:
    # initializam generatorul de numere aleatorii
    random.seed()
    for k in range(n - 1):
        # generam un numar aleatoriu
        # cuprins intre 1 si 1000
        x = random.randint(1, 1000)
        f.write(str(x) + " + ")
    f.write(str(random.randint(1, 1000))))
```

În continuare, vom implementa un program care să calculeze suma numerelor din fișier și apoi să o scrie la sfârșitul primei linii, după un semn "=" (de exemplu, 12 + 300 + 9 + 1234 = 1555):

```
numef = input("Numele fisierului: ")

# deschidem fisierul pentru citire si
# calculam suma numerelor de pe prima linie
f = open(numef, "r")
nr = [int(x) for x in f.readline().split("+")]
s = sum(nr)

# deschidem fisierul pentru adaugare
# si scriem suma la sfarsitul primei linii

# fisierul deschis anterior pentru citire
# va fi inchis implicit
f = open(numef, "a")
f.write(" = " + str(s))
# inchidem explicit fisierul
f.close()
```

# Funcții

O *funcție* este un set de instrucțiuni, grupate sub un nume unic, care efectuează o prelucrare specifică a unor date în momentul în care este apelată.

Funcțiile sunt utilizate pentru a modulariza codul dintr-un program, prin împărțirea sa în subprograme (i.e., funcții) care execută fiecare o singură prelucrare bine definită. În plus, funcțiile permit o reutilizare simplă și sigură a codului, prin organizarea lor în module și pachete.

În limbajul Python sunt predefinite mai multe funcții care sunt des utilizate în programe, cum ar fi funcții de conversie (int, str, bool etc.), funcții matematice (min, max, sum etc.), funcții pentru crearea unei colecții (list, dict, set etc.) ș.a.m.d. O listă completă a tuturor funcțiilor predefinite din limbajul Python poate fi consultată aici: https://docs.python.org/3/library/functions.html.

O funcție poate fi definită de către un utilizator folosind cuvântul cheie def, astfel:

```
def nume_funcție(parametrii formali)
    instrucțiuni
```

*Parametrii formali* sunt utilizați în antetul unei funcții pentru a preciza datele sale de intrare într-un mod abstract (formal). *Parametrii actuali* ai unei funcții sunt expresii ale căror valori sunt asociate parametrilor formali în momentul apelării unei funcții.

Parametrii formali ai unei funcții pot fi de mai multe tipuri în limbajul Python:

 parametri simpli au, de obicei, un caracter pozițional, respectiv în momentul apelării unei funcției trebuie să fie înlocuiți cu același număr de parametrii efectivi, compatibili ca tipuri de date:

```
def suma(x, y):
    return x + y

# apelare pozițională
s = suma(3, 7)
print("s =", s)

# apelare prin nume
s = suma(y=7, x=3)
print("s =", s)
```

În acest exemplu, parametrii formali ai funcției suma sunt x și y, iar constantele 3 și 7 sunt parametrii efectivi. Deoarece în limbajul Python nu se precizează explicit tipurile de date ale parametrilor formali, compatibilitatea dintre ei și parametrii efectivi poate

fi, în unele cazuri, o noțiune destul de vagă. Astfel, în acest exemplu, parametrii efectivi ai funcției pot fi și două șiruri de caractere sau două liste, dar nu pot fi, de exemplu, un număr și un șir de caractere! De asemenea, se observă faptul că parametrii simpli permit și apelarea prin nume.

• parametri cu valori implicite sunt utilizați dacă se dorește inițializarea lor cu niște valori implicite, care vor fi utilizate în momentul apelării funcției cu un număr de parametrii efectivi mai mic decât numărul parametrilor formali:

Atenție, dacă antetul unei funcții conține și parametri simpli și parametri cu valori implicite, parametrii poziționali trebuie să fie precizați primii, pentru a evita ambiguitățile care pot să apară în momentul apelării funcției! De exemplu, dacă funcția se mai sus ar avea antetul suma(x=0, y), atunci apelul s=suma(7) ar putea fi interpretat fie ca s=suma(7,y) și ar fi incorect (parametrului y nu i-ar fi asociat niciun parametru efectiv), fie ca s=suma(0,7) și ar fi corect. În schimb, antetul suma(x,y=0) nu va mai genera nicio ambiguitate în cazul apelului s=suma(7), putând fi interpretat doar ca s=suma(7,0).

În limbajul Python, se pot defini foarte simplu şi funcții cu *număr variabil de parametri*, care sunt utile când nu se poate preciza numărul exact de parametrii ai unei funcții. De exemplu, în același program putem să avem nevoie de o funcție care să calculeze suma a două numere și de o funcție care să calculeze suma a trei numere. Deoarece în limbajul Python nu se pot defini două funcții cu același nume și număr diferit de parametri (de fapt, se pot defini, dar compilatorul o va lua în considerare doar pe ultima definită și orice apelare a primei funcții definite va fi semnalată ca eroare), înseamnă că ar trebui să definim două funcții cu nume diferite, dar care sunt foarte asemănătoare din punct de vedere al prelucrărilor efectuate.

Definirea unei funcții cu număr variabil de parametri se realizează adăugând simbolul \* înaintea unui parametru, ceea ce indică faptul că parametrul respectiv va conține, sub forma unui tuplu, un număr oarecare de parametri efectivi. De exemplu, o funcție cu număr variabil de parametri care calculează suma acestora se poate defini și apela astfel:

```
def suma(*args):
    sa = 0
    for x in args:
        sa = sa + x
    return sa
s = suma()
print("s =", s)  #s = 0

s = suma(7)  #s = 7
print("s =", s)

s = suma(1, 2, 3, 4)  #s = 10
print("s =", s)

lista = [x for x in range(1, 11)]
s = suma(*lista)
print("s =", s)  #s = 55
```

Evident, o funcție poate să aibă un singur parametru variabil, iar eventualii parametrii existenți după el trebuie să fie precizați explicit, prin numele lor, în momentul apelării funcției (deoarece ei nu mai pot fi accesați pozițional, fiind precedați de un număr necunoscut de parametrii). De exemplu, următoarea funcție calculează suma parametrilor variabili care sunt cel puțini egali cu valoarea minim:

Dacă în exemplul de mai sus nu vom preciza explicit valoarea parametrului minim (de exemplu, scriind s = suma(1, 2, 3, 10)), atunci se va semnala o eroare de tipul TypeError (pentru exemplul considerat aceasta va fi TypeError: suma() missing 1 required keyword-only argument: 'minim').

Bineînțeles, dacă o funcție cu număr variabil de parametri are și parametri simpli și parametri cu valori implicite, atunci trebuie respectată și regula ca parametrii simpli să

fie scriși înaintea celor cu valori implicite! De exemplu, următoarea funcție suma calculează suma parametrilor variabili cuprinși între valorile minim și maxim care sunt și multiplii ai numărului t:

```
def suma(t, *valori, minim=0, maxim=100):
    sv = 0
    for x in valori:
        if minim <= x <= maxim and x % t == 0:
            sv = sv + x
    return sv

s = suma(7, 14, 19, 21, -56)  #s = 35

rint("s =", s)  #s = -21

print("s =", s)

s = suma(7, 14, 19, 21, -56, minim=-100)  #s = -21

print("s =", s)

s = suma(7, 14, 19, 21, -56, minim=-100, maxim=0)  #s = -56

print("s =", s)</pre>
```

Datorită operației implicite de împachetare a mai multor valori sub forma unui tuplu, o funcție poate să furnizeze mai multe valori. Astfel, o instrucțiune de forma return a, b este echivalentă cu return (a, b), așa cum se poate observa din următorul exemplu:

```
def suma_prod(x, y):
    return x + y, x * y

s, p = suma_prod(3, 7)
print("s =", s, "\tp =", p)

(s, p) = suma_prod(3, 7)
print("s =", s, "\tp =", p)

t = suma_prod(3, 7)
print("s =", t[0], "\tp =", t[1])
print("Suma si produsul:", *t)
```

Observați faptul că la apelarea funcției se poate utiliza operația de despachetare a unui tuplu, complementară celei de împachetare utilizată în instrucțiunea return!

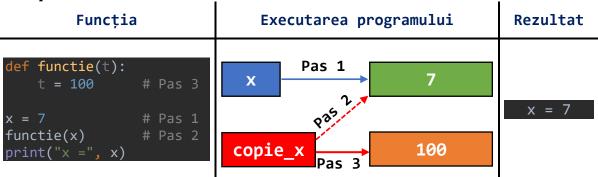
# Modalități de transmitere a parametrilor

În limbajele C/C++ transmiterea unui parametru efectiv către o funcție se poate realiza în două moduri:

- transmitere prin valoare: se transmite o copie a valorii parametrului efectiv, deci modificările efectuate asupra parametrului respectiv în interiorul funcției NU se reflectă și în exteriorul său;
- transmitere prin adresă/referință: se transmite adresa parametrului efectiv, deci modificările efectuate asupra parametrului respectiv în interiorul funcției se reflectă și în exteriorul său.

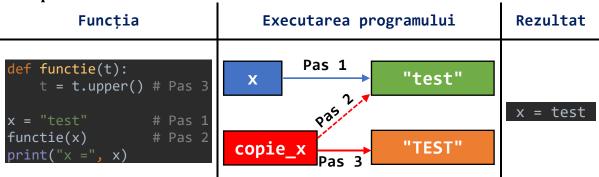
În limbajul Python, orice parametru efectiv al unei funcții este o referință către un obiect (i.e., nu se transmite obiectul în sine, ci o referință spre el) care se transmite implicit prin valoare (i.e., se transmite o copie a referinței respective), deci modificarea referinței respective în interiorul funcției nu se va reflecta în exteriorul său. Din acest motiv, mecanismul de transmitere a parametrilor către o funcție în limbajul Python se numește transmitere prin referință la obiect (call by object reference).

# Exemplul 1:



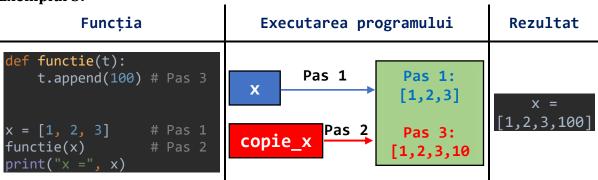
**Explicație:** După pasul 2, variabilele x și copie\_x vor conține aceeași referință (spre obiectul 7), dar după pasul 3 doar copie\_x se va modifica (va conține referința obiectului 100), deoarece x este o referintă transmisă prin valoare.

## **Exemplul 2:**



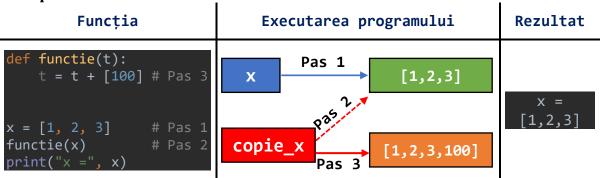
**Explicație:** vezi exemplul anterior!

## Exemplul 3:



**Explicație:** După pasul 2, variabilele x și copie\_x vor conține aceeași referință, iar după pasul 3 conținutul listei se va modifica prin intermediul referinței din copie\_x (însă fără a modifica referința listei!), deci variabila x va putea accesa lista modificată.

## Exemplul 4:



**Explicație:** După pasul 2, variabilele x și copie\_x vor conține aceeași referință (spre lista [1,2,3]), dar la pasul 3 operatorul de concatenare (+) va crea o nouă listă [1,2,3,100], deci după pasul 3 doar copie\_x va conține referința noii liste. Practic, deși o listă este mutabilă, elementul 100 a fost adăugat într-o manieră specifică obiectelor imutabile!

În concluzie, mecanismul de *transmitere prin referință la obiect* a parametrilor efectivi către o funcție în limbajul Python utilizează transmiterea prin valoare a referințelor parametrilor efectivi (copii ale referințelor lor, deci modificarea acestora nu va fi vizibilă în exteriorul funcției) și acționează astfel:

- dacă obiectul asociat referinței este imutabil, atunci modificarea parametrului respectiv în interiorul funcției nu se va reflecta în exteriorul funcției deoarece conținutului obiectului asociat referinței nu poate fi modificat (din cauza imutabilității), iar crearea unei referințe noi nu se va reflecta în exteriorul funcției (din cauza transmiterii prin valoare a referinței);
- dacă obiectul asociat referinței este *mutabil*, atunci modificarea conținutului parametrului respectiv în interiorul funcției se va reflecta în exteriorul funcției deoarece nu se va modifica referința obiectului.

Atenție la exemplul 4 de mai sus, deoarece acolo nu se modifică direct conținutul obiectului mutabil de tip listă, ci se creează un nou obiect (tot o listă mutabilă) care conține noua listă și se atribuie referința sa copiei referinței parametrului!